



TAMPEREEN TEKNILLINEN YLIOPISTO

**SAMI PEURALA**  
**LEVYTYÖKONEEN LAADUN TARKKAILU VIRTAMITTAUSTEN**  
**ANALYSOINNILLA**  
Diplomityö

Tarkastaja: professori Tarmo Lipping  
Tarkastaja ja aihe hyväksytty 29. loka-  
kuuta 2018

# TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO

Tietotekniikan koulutusohjelma

**PEURALA, SAMI:** Levytyökoneen laadun tarkkailu virtamittausten analysoinnilla

Diplomityö, 43 sivua

Marraskuu 2018

Pääaine: Ohjelmistotuotanto ja tiedonhallinta

Tarkastaja: professori Tarmo Lipping

Ohjaaja: yliopistonlehtori Jari Turunen

Avainsanat: virtamittaus, analyysi, aikasarja, luokittelu, levytyökone

Levytyökoneen pääakseleista mitataan levytyökoneen valmistuksen loppuvaiheessa virtamittaukset. Virtamittauksia on analysoitu visuaalisesti, mutta kaikkia valmistuvia levytyökoneita ei ehditä tarkastamaan virtamittausten osalta. Tässä diplomityössä on pyritty löytämään analysointimenetelmiä ja koneellisia luokittelumenetelmiä tunnistamaan poikkeavat virtamittaukset.

Diplomityön aluksi käydään läpi eri menetelmiä, joiden avulla voitaisiin kehittää koneellinen analyysi ja luokittelumenetelmä tunnistamaan poikkeavat virtamittaukset. Tässä työssä tutkimuskohteeksi valittiin aikasarjan analyysimenetelmiä ja aikasarja-analyysin perusteella valittiin parhaimmat piirteet koneelliseen luokitteluun. Erilaisia luokittelumenetelmiä testattiin useilla eri levytyökonetyyppien virtamittauksilla. Monipuolinen aineisto mahdollistaa huomioimaan eri levytyökonetyyppien pääakseleiden eroavaisuudet virtamittausten osalta ja kehittämään luokittelumenetelmän siten, että kaikkien eri levytyökonetyyppien luokittelu on mahdollisimman luotettava.

Levytyökoneiden pääakseleiden virtamittauksia analysointiin Python-ohjelmointikielellä, johon on saatavilla ohjelmakirjastoja aikasarjan analysointiin sekä koneellisen luokittelun toteuttamiseen. Toteutettu luokittelumenetelmä pystyy luokittelemaan eri levytyökoneiden pääakseleiden virtamittauksia kolmeen eri luokkaan.

Toteutettu luokittelumenetelmä pystyi tunnistamaan luotettavasti toimenpiteitä vaativat poikkeavat virtamittausarvot. Diplomityö osoitti, että koneelliseen analysointiin ja luokitteluun pohjautuva järjestelmä pystyy tunnistamaan toimenpiteitä vaativat poikkeavat virtamittausarvot. Menetelmää voidaan soveltaa myös siten, että jokaiselle eri levytyökonetyypille kehitetään oma koneellinen analysointi- ja luokittelumenetelmänsä.

# ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme In Information Technology

**PEURALA, SAMI:** Quality assessment of sheet metal machines by measuring electrical current

Master of Science Thesis, 43 pages

November 2018

Major: Software Engineering and Information Management

Examiner: Professor Tarmo Lipping

Instructor: University lecturer Jari Turunen

Key words: electrical current measurements, analysis, time-series, classification, sheet metal machine

Electrical current measures are taken from sheet metal machine main axes at the final stage of manufacturing. Electrical current measurements were investigated via visual inspection, but it was not possible to handle all the machines in this way. In this Master's thesis it has been pursued to develop a machine based analysis and classification method to recognize abnormal electrical current measurements.

At the beginning of this Master's thesis I go through different methods helpful in developing an automated analysis and classification method for detecting abnormal electrical current measurements. Time-series analysis methods were selected to get features from electrical current measurements and best features were selected for classification. Different kinds of machine based classification methods were analysed with electrical current measurements of several sheet metal machines. Various sheet metal machine's main axes types were taken into account with versatile electrical current measurements. Versatile electrical current measurements help to development analysis and classification methods that take into account all sheet metal machine main axes types.

Python programming language was used to implement the analysis and classification algorithms. Python programming environment can be extended via time series analysis libraries and classification libraries. The implemented classification method can classify electrical current measurements into three different classes.

This work shows that automated analysis and classification system can detect abnormal electrical current measurements. The implemented method can be adapted to use different analysis and classification methods for each sheet metal machine type.

## ALKUSANAT

Tämä diplomityö on tehty Finn-Power Oy:n Global Service yksikölle kehittämään levytyökoneiden laadun valvontaa virtamittausten analysoinnilla.

Haluan esittää kiitokset työn tarkastaja toimineelle professori Tarmo Lipping:lle arvokkaista neuvoista sekä tuesta. Osoitan kiitokset myös työn ohjaajana toimineelle yliopistonlehtori Jari Turuselle saamistani ohjeista ja neuvoista.

Haluan myös kiittää kaikkia lähipiirissäni olevia henkilöitä saamastani tuesta sekä kannustuksesta saattaa aloittamani opiskelu päätökseen.

Seinäjoella 19.11.2018

Sami Peurala

# SISÄLLYS

TIIVISTELMÄ.....	I
ABSTRACT.....	II
ALKUSANAT.....	III
1 JOHDANTO.....	1
2 LEVYTYÖKONEIDEN PÄÄAKSELEIDEN VIRTAMITTAUKSET.....	3
2.1 Erilaiset levytyökonemallit.....	3
2.2 Nykyaikaisen levytyökoneen moottorityypit.....	3
2.3 Virtamittausohjelma.....	4
2.4 Virtamittaustiedon siirtäminen.....	4
3 VIRTAMITTAUSTIEDON ANALYSOINNIN MENETELMIÄ.....	6
3.1 Aikasarja.....	6
3.2 Poikkeavat arvot.....	6
3.3 Liukuva keskiarvo.....	7
3.4 Histogrammi.....	8
3.5 Huipukkuus ja vinous.....	9
3.6 Korrelaatio.....	11
3.7 Entropia.....	12
3.8 Keskihajonta.....	13
3.9 Diskreetti Fourier-muunnos.....	13
3.10 Tehotiheysspektri, Welch'in menetelmä.....	14
4 VIRTAMITTAUSTIETOJEN LUOKITTELUN MENETELMIÄ.....	16
4.1 Piirteiden määrä ja valinta.....	16
4.2 K-lähimmän naapurin luokittelija.....	17
4.3 Tukivektorikone ( Support Vector Machine, SVM).....	18
4.4 Päättöspuu-luokittelija.....	20
4.5 Lineaarinen diskriminanttianalyysi (LDA).....	23
4.5.1 Neliöllinen diskriminanttianalyysi (QDA).....	24
5 VIRTAMITTAUSTIETOJEN ANALYSOINNIN TULOKSET.....	26
5.1 Levytyökoneiden virtamittausten koostaminen.....	26
5.2 Raja-arvojen määrittely.....	29
5.3 Piirteiden valinta.....	33
5.4 Virtamittaustietojen luokittelu.....	36
5.5 Luokittelun arviointi.....	37
5.6 Levytyökoneen pääakselin vikaantumisen ennustaminen.....	39
5.7 Eri levytyökonemallien vastaavuudet.....	39
6 JOHTOPÄÄTÖKSET.....	42
6.1 Yhteenveto.....	42
6.2 Sovellettavuus.....	43

6.3	Jatkotoimenpiteet.....	43
LÄHTEET.....		44

# 1 JOHDANTO

Levytyökoneista kerätään ja mitataan erilaisia tietoja niiden tuotantokäytön aikana. Levytyökoneesta kerätty tieto talletettiin vuoteen 2010 asti paikalliselle ohjaustietokoneelle. Kerätyistä tiedoista näki mittausten ja muun kerätyn tiedon kumulatiivisen kertymän. Kerätyn ja mitatun tiedon kumulatiivista tallennusta käytettiin, koska ohjaustietokoneella olevat resurssit olivat rajallisemmat kuin nykyisissä ohjaustietokoneissa. Kumulatiivisessa tallennuksessa reaaliaikaohjaukselle kerätty tieto summattiin aikaisemman tiedon kanssa paikalliseen tietokantaan edellisen tiedon kanssa. Tällä tavoin kerätyn tiedon historia menetettiin, mutta kokonaismäärät saatiin talletettua.

Aikoinaan tällainen kumulatiivinen mittausten ja tietojen tallentaminen paikalliselle ohjaustietokoneelle oli riittävä siihen, että levytyökoneen käyttömäärää voitiin seurata. Paikallisesti talletettujen kumulatiivisten tietojen perusteella voitiin tarkastella esimerkiksi yksittäisten akseleiden ajomatkoja. Levytyökoneen määräaikaishuollossa voitiin tarkastella kerättyjä tietoja, mutta niiden perusteella ei pystytty tekemään tarkempaa analyysia jonkin yksittäisen osan kulumisesta.

Internetin ja tiedonsiirtolaitteiden kehittyessä kerättyjä tietoja voitiin siirtää yksittäisestä levytyökoneesta koneen valmistajalle. Tämä mahdollisti kerättyjen ja mitattujen tietojen tarkastelemisen myös muuna aikana kuin määräaikaishuoltojen yhteydessä. Tietojen siirtäminen yksittäiseltä levytyökoneelta ei ollut automaattista ja usein kerättyjä tietoja ladattiin silloin, kun jokin muun huolto tai päivitys tuli ajankohtaiseksi levytyökoneelle.

Ohjaustietokoneiden kapasiteetin lisääntyessä levytyökoneen ohjaukselta voitiin mitata ja kerätä aiempaa enemmän tietoja ja tallettaa niitä ohjaustietokoneelle. Lisääntyneen tiedon tallentamiseen käytettiin edelleen kumulatiivisia menetelmiä, koska valmistajaa ja asiakkaita kiinnosti erityisesti eri mittausten ja kerätyn tiedon kokonaismäärät.

Internetin palvelujen ja digitalisaation kehittyessä teollinen internet alkoi muodostumaan. Teollisen internetin ja esineiden internetin yleistyessä niihin liittyvien palveluiden ja laitteiden tarjonta lisääntyi merkittävästi. Teollisen internetin ja pilvipalvelujen kehittymisen johdosta tuli mahdolliseksi tallettaa kerättyä ja mitattua tietoa siten, että tiedon historia säilytetään. Teollisen internetin laitteet mahdollistivat salatun ja turvallisen tavan siirtää automaattisesti mitattua ja kerättyä tietoa levytyökoneista ympäri maailman.

Tuotantovaatimusten kasvaessa levytyökoneille tehtävät määräaikaishuoltojen ajoittamiset tulevat yhä tärkeimmiksi. Tuotantovaatimusten kasvaessa myös tarvittavien huoltojen ennakointi tulee yhä tärkeämmäksi tulevaisuudessa. Nykyaikaiset tuotantoprosessit ovat haavoittuvaisempia ennakoimattomille tuotantokatkoksille ja aiheuttavat isompia ongelmia koko toimitusketjussa. Huoltojen ajoittaminen ja ennakointi lisäävät tuotannon tehokkuutta ja näistä tekijöistä johtuen levytyökoneesta mitatun ja kerätyn tiedon analysointiin ja tutkimiseen tarvitaan kehitystä.

Tässä diplomityössä analysoidaan ja tutkitaan jo olemassa olevaa tietoa levytyökoneista huollon tarpeiden perusteella. Analysoinnissa ja tutkimuksessa pyritään selvittämään ja saamaan vastauksia seuraaviin kysymyksiin:

- eri levytyökonemallien pääakseleiden ohjausvirran raja-arvot ja niiden määrittäminen.
- yksittäisen akselin vikaantumisen tunnistaminen automaattisesti.
- yksittäisen akselin vikaantumisen ennustaminen.
- eri levytyökonemallien yhteneväisyydet rakenteen ja kerätyn tiedon perusteella.

Jokainen yllä oleva kohta on itsessään hyvin laaja kokonaisuus. Tässä tutkimuksessa ja analyysissä pyritään kehittämään ensimmäiset arviot sekä vastaukset yllä olevien kohtien pohjaksi. Huollon tarpeet huomioon ottavaa analyysia ja tutkimusta kerätyn ja mitatun tiedon perusteella on tarkoitus jatkaa sekä laajentaa tämän tutkimuksen ja analyysin pohjalta.

Luvussa kaksi kerrotaan levytyökoneen virtamittauksen periaate ja virtamittauksien siirto internetin yli valmistajan pilvipalveluun. Luvussa kolme kuvataan virtamittausten analysoinnin eri menetelmiä. Luvussa neljä tarkastellaan virtamittausten luokittelun eri menetelmiä. Luvussa viisi esitellään analysoinnin ja luokittelun tuloksia ja luvussa kuusi kerrotaan syntyneet johtopäätökset ja jatkotoimenpiteet.



## **2 LEVYTYÖKONEIDEN PÄÄAKSELEIDEN VIRTAMITTAUKSET**

Nykyaikaisista levytyökoneista mitataan pääakseleiden virta-arvoja säännöllisen ajanjakson mukaan. Lävistävässä levytyökoneissa pääakseleihin katsotaan kuuluvan x-, y- ja ram-akseli. Nykyaikaiseen levytyökoneeseen voidaan liittää kappale- ja materiaaliautomaatioon liittyviä laitteita, kuten lastaus- ja pinontalaitteita. Levytyökoneeseen liitettyjen laitteiden mitatut virta-arvot talletetaan myös levytyökoneen ohjaustietokoneelle, mutta tässä työssä analysoidaan levytyökoneen pääakseleiden virta-arvoja.

### **2.1 Erilaiset levytyökonemallit**

Levytyökonemallit voidaan jakaa karkeasti kahteen eri luokkaan, lävistäviin ja leikkaaviin koneisiin. Lävistävässä levytyökoneissa käytetään lävistäviä työkaluja, joilla lävistetään ohutlevymateriaalia. Leikkaavat levytyökoneet käyttävät pääsääntöisesti ohutlevyn leikkaamiseen vettä, plasmaa tai laseria. Yhdistelmä-levytyökoneissa on yhdistetty lävistys ja laserleikkaus.

Tässä työssä analysoidaan kolmen eri levytyökonemallin pääakseleiden virta-arvoja. Analysoitavat levytyökonemallit ovat lävistävä levytyökone kahdella eri teknisellä ratkaisulla toteutettuna sekä yhdistelmä-levytyökone. Kaikissa kolmessa levytyökonemallissa on samankaltainen rakenne ohutlevyn liikuttamiseen levytyökoneen työalueella. Ohutlevyn liikuttaminen perustuu kelkkaan, johon työstettävä ohutlevy kiinnitetään. Kelkka liikkuu x- ja y-suunnassa, jotka muodostavat levytyökoneen pääakselit x ja y. Levytyökoneen rungossa oleva lävistysakseli, eli ram-akseli on levytyökoneen kolmas pääakseli.

### **2.2 Nykyaikaisen levytyökoneen moottorityypit**

Nykyaikaisissa levytyökoneissa on käytössä kahdentyyppisiä moottoreita: servomoottoreita ja lineaarimoottoreita. Servomoottorien kanssa käytetään mekaanisia välityksiä, jonka avulla moottorin pyörimisliike saadaan muutettua suoraviivaiseksi liikkeeksi. Servomoottorit ovat tyypillisesti kiinnitettynä kuularuuviin, jonka avulla servomoottorin pyörimisliike saadaan muutettua suoraviivaiseksi liikkeeksi. Lineaarimoottori on tyypillisesti kiinnitettynä suoraan johteisiin, joka saa aikaan suoraviivaisen liikkeen. Li-

neaarimoottoreissa ei tarvita mekaanisia välityksiä suoraviivaisen liikkeen aikaansaamiseksi.

Servomoottoria ohjataan elektronisen servo-ohjaimen avulla. Servomoottoreissa voi olla myös sisäänrakennettuna pulssianturi, jolla voidaan mitata akselin kulmaa ja pyörimiskertoja. Lineaarimoottoria ohjataan samaan tapaan kuin servomoottoria eli erillisen ohjaimen avulla.

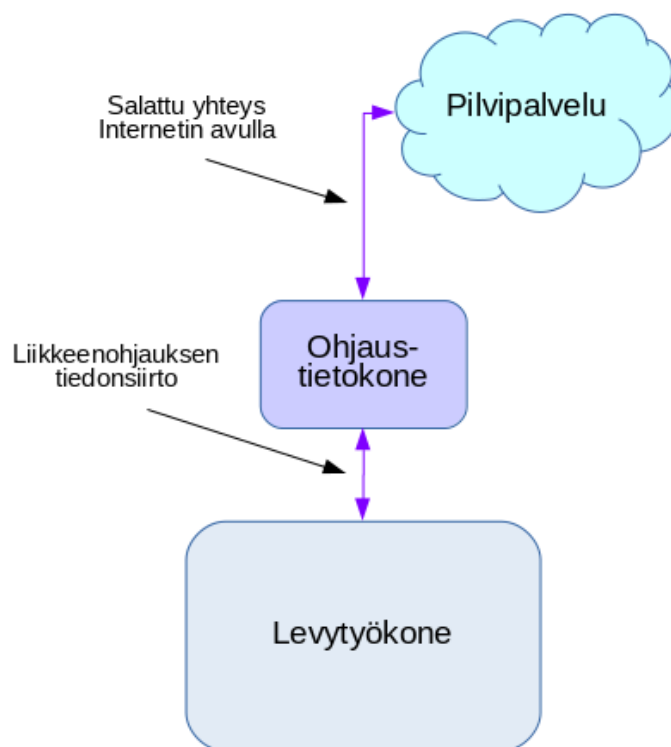
## 2.3 Virtamittausohjelma

Nykyaikaisen levytyökoneen ohjauksena on numeeriseen ohjaukseen ja ohjelmoitavaan logiikkaan perustuva ohjaus. Levytyökoneen ohjaustietokone suorittaa sekä reaaliaikaisen ohjauksen ajon, että levytyökoneen käyttöliittymän ajon. Levytyökoneella ajetaan erillinen virtamittausohjelma, jonka tulokset talletetaan levytyökoneen ohjaustietokoneelle. Ensimmäiset levytyökoneen toimittajan järjestelmään talletetut virtamittauksien tulokset saadaan, kun levytyökone on hyväksytysti käyntiin ajettu valmistajan tuotannossa. Seuraavat pääakselien virtamittaukset tehdään asiakkaan tuotannon lomassa ajamalla erillinen virtamittausohjelma.

Levytyökoneiden virtamittaukset tehdään ajamalla akseleita vakio nopeudella ja samalla mitataan tarvittava virta akseleiden moottoreilta. Jokainen levytyökoneen pääakseli mitataan erikseen. Vakionopeudella tehtävä virtamittaus kertoo akselin eri kohdissa tarvittavan virta-arvon. Eri konemalleista saadaan erilaiset mittauksien tiedot pääakseleilta, koska niiden tekninen toteutus on erilainen. Virtamittaukset tehdään jokaisen pääakselin koko pituuden osalta. Näin saadaan koordinaatteihin perustuva aikasarja virtamittauksesta koko akselin pituudelta. Virtamittauksia suoritetaan säännöllisen ajanjakson välein. Näin saadaan talletettua tietoa, josta voidaan saada informaatiota koneen pääakselien kunnosta ja mahdollisesta kulumisesta.

## 2.4 Virtamittauksien tiedon siirtäminen

Levytyökoneen toimittajan järjestelmä noutaa virtamittauksien tietoja ohjaustietokoneelta säännöllisin välein ja haetut tulokset talletetaan levytyökoneen toimittajan pilvipalveluun myöhemmää tarkastelua varten. Virtamittauksien tietojen haku tehdään salatun yhteyden avulla julkisen internetin kautta. Salattu yhteys varmistaa tietoturvallisen tiedonsiirron ohjaustietokoneen ja levytyökoneen valmistajan järjestelmän välillä. Kuvassa 1 on kuvattuna virtamittauksien tiedon noudon periaate.



**Kuva 1.** Virtamittaustietojen noudon periaate.

### 3 VIRTAMITTAUSTIEDON ANALYSOINNIN MENETELMIÄ

Aikasarjatiedon käsittelyyn ja analysointiin on olemassa erilaisia menetelmiä. Tavoitteena on löytää sopivat menetelmät levytyökoneiden virtamittausten analysointiin. Erilaiset analysointitavoitteet asettavat erilaiset vaatimukset analyysimenetelmille. Alla olevissa kappaleissa on pohdittu erilaisten aikasarjan käsittelymenetelmien sopivuutta levytyökoneesta saatavan virtamittaustiedon analysointiin.

#### 3.1 Aikasarja

Aikasarja määritellään havaintojen jonoksi, jossa aika määrittelee havaintojen järjestyksen. Yleisesti oletetaan, että yksittäiset havainnot saavat kvantisoituja arvoja. Havainnot talletetaan määrätyn aikavälein ja näin saadaan aikaan perustuva diskreetti aikasarja. Aikasarja voidaan määritellä:  $x_i, i=1,2,\dots,n$ , jossa ajan  $i$  hetket määrittävät havaintoperiodin pituuden jolta havainnot on kerätty. Aikasarjoja kuvataan yleisesti graafisesti ajan funktiona [1].

#### 3.2 Poikkeavat arvot

Levytyökoneen virtamittauksessa saattaa esiintyä eri syistä johtuen poikkeavia arvoja. Poikkeavien arvojen tunnistaminen ja poistaminen ennen analysointia lisää analyysin tarkkuutta. Jokaisella eri levytyökonetyypillä on mekaanisen rakenteen asettamat rajat pääakseleiden osalta, jotka määrittävät pääakseleiden virtamittausten pituuden. Levytyökoneista mitatut pääakseleiden virta-arvot voivat sisältää poikkeavia arvoja monenlaisista syistä johtuen. Poikkeavia arvoja voivat olla myös puuttuvat arvot, jotka voivat johtua esimerkiksi kesken jääneestä virtamittausprosessista. Poikkeavia arvoja levytyökoneissa saattavat olla myös liian pitkät virtamittausarvot, jotka ovat syntyneet virtamittausprosessin häiriön aikana. Levytyökoneista saatavat arvot voivat vaikuttaa poikkeavilta, mutta poikkeavien arvojen aiheuttajaksi voi osoittautua mekaaninen vika, kuluma tai ongelma.

Mekaanisen vian, kuluman tai ongelman tunnistaminen havaintoarvojen perusteella on haasteellista ja tällaisia havaintoarvoja ei saa automaattisesti pitää poikkeavina arvoina. Poikkeavat arvot tulisi pystyä tunnistamaan siten, että niiden aiheuttaja voidaan

määrittää, joko mekaaniseksi viaksi, kulumaksi tai muuksi ongelmaksi itse mittausprosessin sijaan. Mekaanisen vian, kuluman tai ongelman aiheuttamat poikkeamat eivät muuta pääakselin virtamittauksen pituutta, mutta aiheuttavat merkittäviä poikkeamia virtamittausarvoihin. Tyypillisesti mekaanisen vian, kuluman tai ongelman aiheuttamat poikkeamat näkyvät virtamittauksen alussa ja lopussa merkittävänä poikkeamina virta-arvoissa. Osa mekaanisen vian, kuluman tai ongelman aiheuttamista poikkeavista virtamittausarvoista voivat näkyä virtamittauksen epäsäännöllisinä muutoksina koko pääakselin pituuden osalta.

### 3.3 Liukuva keskiarvo

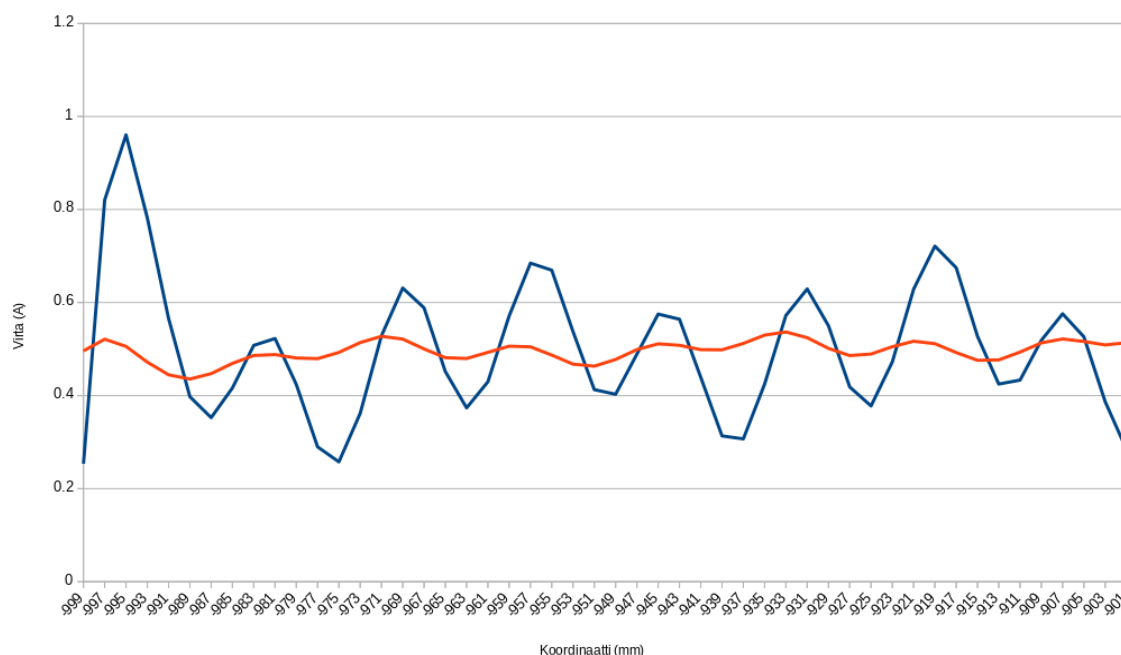
Yksinkertaisessa liukuvassa keskiarvossa ei käytetä painokertoimia arvoille. Liukuvaa keskiarvoa voidaan käyttää tasoittamaan lyhytaikaista vaihtelua. Liukuvan keskiarvon avulla saadaan trendi ja pitkäaikainen vaihtelu esille. Keskiarvo lasketaan seuraavasti:

$$\frac{1}{n} \sum_{i=1}^n x_i \quad (3-1)$$

Painotettu keskiarvo lasketaan seuraavasti:

$$y(t) = \frac{1}{n} \sum_{i=0}^{n-1} m_i x_{t+i} \quad (3-2)$$

missä  $n$  määrittelee liukuvan ikkunan koon,  $m_i$  määrittelee käytettävät painokertoimet ja  $x_t$  määrittelee havaintoarvon [2]. Kuvassa 2 on havainnollistettu liukuvan ikkunan avulla tasoitettuja havaintoarvoja. Kuvassa 2 on sinisellä värillä kuvattuna alkupe-  
räiset havaintoarvot ja punaisella liukuvalla ikkunalla tasoitetut havaintoarvot. Kuvassa 2 olevassa esimerkissä on käytetty painokertoimina ykköstä, jolloin kaikki arvot saavat saman painotuksen.



**Kuva 2.** Liukuvalla ikkunalla tasoitetut havaintoarvot.

Liukuvan keskiarvon avulla saataneen paremmin levytyökoneen pääakselin virtakäyrän muoto esiin. Pääakselin virtakäyrän muodosta voidaan päätellä miten virta-arvot jakautuvat pääakselin pituuden osalta. Tällä tavoin voidaan paremmin tutkia ja analysoida pääakselin eri kohdissa olevia muutoksia.

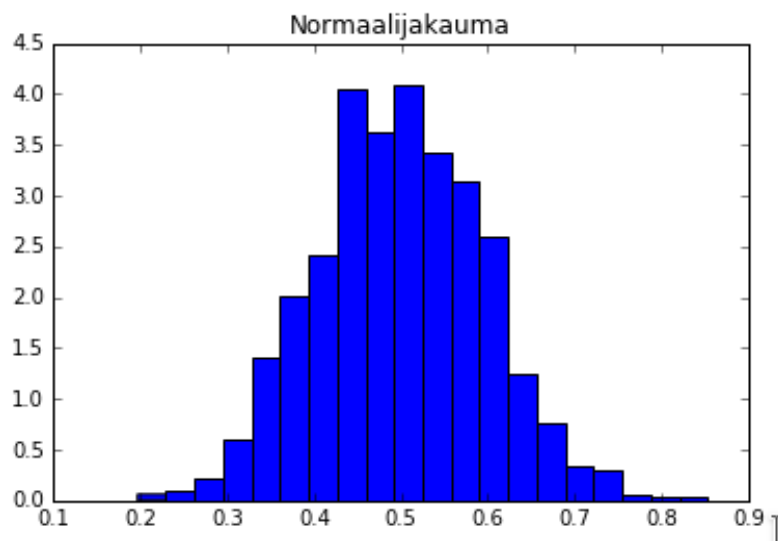
### 3.4 Histogrammi

Histogrammissa havaintoarvot jaetaan samansuuruisiin alueisiin ja jokainen havaintoarvo sijoitetaan arvonsa perusteella valittuun alueeseen. Histogrammin alueet voidaan jakaa tasaisiin alueisiin esimerkiksi havaintoarvojen lukumäärän perusteella. Histogrammi kuvaa havaintoarvojen jakauman ja sen perusteella voidaan arvioida jakauman tyyppiä. Histogrammit piirretään pylväsmuodossa, jossa pylvään leveys kuvaa alueen väliä ja pylvään korkeus kertoo montako arvoa alueeseen kuuluu. Lukumäärästä käytetään nimitystä frekvenssi, kun alueet ovat tasavälisiä. Histogrammin arvot voidaan myös normalisoida siten, että kaikkien alueiden frekvenssien summa on yksi [5]. Histogrammin avulla voidaan arvioida seuraavia mittausarvojen ominaisuuksia [6]:

- mikä on havaintoarvojen jakauma?
- mikä on havaintoarvojen jakauman muoto, symmetrinen vai vinoutunut?

- onko havaintoarvojen joukossa poikkeavia arvoja?

Kuvassa 3 on ohjelmallisesti tuotetun normaalijakautuneen satunnaismuuttujan histogrammi.



**Kuva 3.** Ohjelmallisesti tuotetun normaalijakautuneen satunnaismuuttujan histogrammi.

Levytyökoneen pääakseleiden virtamittausarvojen histogrammikuvaajien avulla voidaan analysoida virta-arvojen jakaumaa, muotoa ja poikkeavia arvoja.

### 3.5 Huipukkuus ja vinous

Huipukkuus ja vinous ovat datanäytteiden amplitudijakauman muotoa kuvaavia tilastotieteellisiä käsitteitä. Normaalijakauman huipukkuus arvo on 3 ja sitä pienempi huipukkuusarvo ilmaisee sen, että jakaumassa on vähemmän poikkeavia arvoja kuin normaalijakaumassa. Huipukkuus lasketaan seuraavasti:

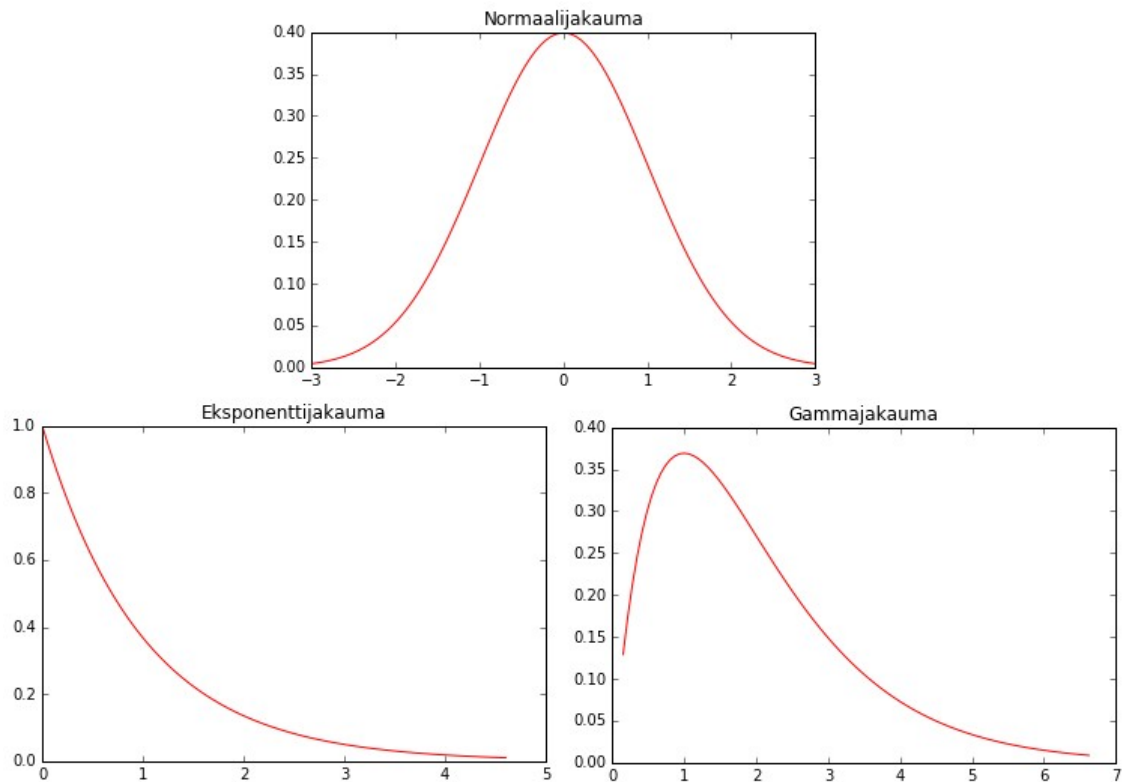
$$\frac{\sum_{i=1}^n (x_i - \bar{x})^4 / n}{s^4} \quad (3-3)$$

missä  $\bar{x}$  on havaintoarvojen keskiarvo,  $s$  on keskihajonta ja  $n$  on havaintoarvojen lukumäärä.

Negatiivinen vinousarvo ilmaisee sen, että yli keskiarvon olevia arvoja on enemmän kuin alle keskiarvon olevia. Positiivinen vinousarvo ilmaisee sen, että alle keskiarvon olevia arvoja on enemmän kuin yli keskiarvon. Vinous lasketaan seuraavasti:

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^3 / n}{s^3} \quad (3-4)$$

Kuvassa 4 on normaalijakauman, gammajakauman ja eksponenttijakauman kuvaajat. Kuvassa 4 olevan normaalijakauman huipukkuusarvoksi saatiin 0,43 ja vinousarvoksi 0,39. Eksponenttijakauman huipukkuusarvoksi saatiin 6 ja vinousarvoksi 2. Gammajakauman huipukkuusarvoksi saatiin 3 ja vinousarvoksi 1,42.



**Kuva 4.** Normaalijakauman, eksponenttijakauman ja gammajakauman kuvaajat.



Esimerkiksi Python-ohjelmointikielessä huipukkuus lasketaan seuraavalla funktiokutsulla:

```
scipy.stats.kurtosis(a, axis=0, fisher=True, bias=True, nan_policy='propagate')
```

johon voidaan määritellä käytetäänkö Person:in vai Fisher:in huipukkuuden määritelmää. Pearson:in määritelmän mukaan normaalijakauman huipukkuusarvo on kolme ja Fisher:in määritelmän mukaan normaalijakauman huipukkuusarvo on nolla.

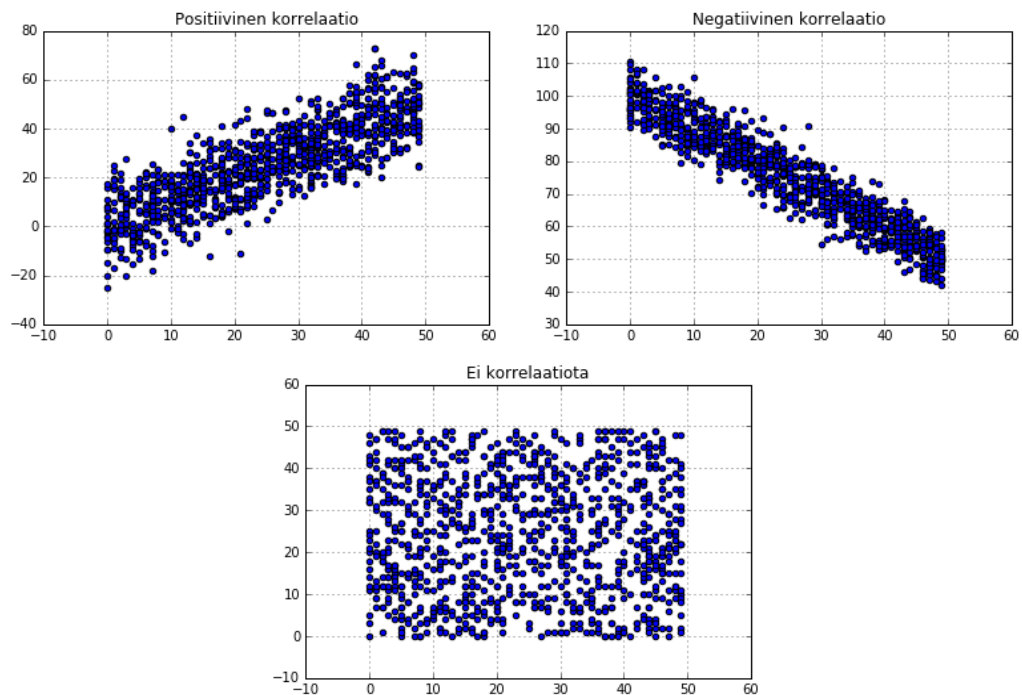
Levytyökoneen pääakseleiden virtamittauksien analysoinnin huipukkuus- ja vinousarvoja voitaneen käyttää virta-arvojen jakauman analysointiin. Pääakseleiden virta-arvojen oletetaan olevan pääosin keskiarvon lähettynneille sijoittuneita. Huipukkuuden ja vinouden arvot levytyökoneen pääakseleiden virta-arvoille voivat olla yksi mahdollisista menetelmistä, jolla voidaan analysoida levytyökoneen pääakseleiden kuluneisuutta tai mekaanisen sovituksen hyvyttä.

### 3.6 Korrelaatio

Korrelaatio kuvaa kahden muuttujan välistä riippuvuutta tilastotieteessä ja todennäköisyyslaskennassa. Havaintoarvoista laskettu korrelaatiokerroin ilmaisee havaintoarvojen välistä korrelaatiota. Korrelaatiota ei ole riippumattomien muuttujien välillä. Korrelaatiokerroin tarkoittaa yleensä Pearsonin korrelaatiokerrointa. Korrelaatiokerroin ei tarkoita kausaalisuutta, eli korrelaatiokerroin ei ole suoraan syy-seuraussuhde. Pearsonin korrelaatiokerroin kertoo kahden muuttujan korrelaation välillä -1:stä 1:een [2].

Korrelaatiota voitaneen käyttää eri levytyökone mallien vertailemiseen keskenään. Eri levytyökoneille on osassa käytetty samaa mekaanista rakennetta joidenkin pääakseleiden osalta. Oletuksena on, että näiden havaintoarvojen pitäisi olla vahvasti korreloituneita. Korrelaatiokerrointa voitaneen käyttää vertailemaan hyvän referenssiarvon turvin jonkin pääakselin havaintoarvojen muutoksia ajan kuluessa.

Kuvassa 5 on kuvaajat positiivisesta korrelaatiosta, negatiivisesta korrelaatiosta ja korreloimattomuudesta.



**Kuva 5.** Positiivisen korrelaation, negatiivisen korrelaation ja korreloimattomuuden kuvaajat.

### 3.7 Entropia

Entropia ilmaisee termodynamiikassa systeemin epäjärjestystä. Yleistettynä entropia kuvaa epäjärjestystä tai kaaosta. Entropiaa voidaan käyttää ilmaisemaan aikasarjan ennustettavuutta. Entropia on suurempi aikasarjassa, jossa on vähemmän säännöllisyyttä tai ennustettavuutta. Informaatioteoriassa entropia ilmaisee viestin tai sanoman informaation määrää [9].

Python-ohjelmointikielessä entropia voidaan laskea seuraavalla funktiokutsulla:

```
scipy.stats.entropy(pk, qk=None, base=None)
```

missä muuttuja *pk* käsittää aikasarjan tai signaalin arvojen todennäköisyyksiä, mistä entropia lasketaan. Entropia lasketaan kaavalla:

$$Z = -\sum p_i \log(p_i) \quad (3-5)$$

Jos entropia lasketaan suhteessa referenssimuuttujaan  $q_k$ , sen laskentaan käytetään kaavaa:

$$Z = \sum p_i \log(p_i/q_i) \quad (3-6)$$

Usein referenssimuuttujaksi valitaan normaalijakautunut muuttuja.

### 3.8 Keskihajonta

Keskihajonta ilmaisee, miten arvot ovat hajautuneet odotusarvon tai keskiarvon ympärille. Keskihajonta saadaan, kun otetaan varianssista neliöjuuri. Keskihajonnan yksikkönä on sama kuin arvojen yksikkö. Otosvariassi lasketaan kaavalla [5]:

$$s^2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n-1} \quad (3-7)$$

missä  $s^2$  on otosvariassi,  $x_i$  on havaintoarvo,  $n$  on arvojen lukumäärä ja  $\bar{x}$  on havaintoarvojen keskiarvo.

Keskihajontaa voitaneen käyttää yhtenä menetelmänä arvioitaessa virtamittauksen muodon vaihtelevuutta. Suurempi keskihajonta voi olla seurausta virtamittauskäyrän amplitudin vaihtelevuudesta.

### 3.9 Diskreetti Fourier-muunnos

Fourier-muunnoksella saadaan aikasarja tai signaali esitettyä taajuuden funktiona. Diskreetti Fourier-muunnos voidaan laskea, jos havainnot ovat tasavälisiä. Jatkuvan funktion  $f(t)$  Fourier-muunnos määritellään integraalina [12]:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j2\pi t f} dt \quad (3-8)$$

missä funktio  $F(\omega)$  ilmaisee aikasarjan tai signaalin taajuuden funktiona. Diskreetti Fourier-muunnos määritellään seuraavasti:

$$F(k) = \frac{1}{n} \sum_{i=0}^{n-1} x_i e^{-j2\pi i k/n}, k=0,1,\dots,n-1 \quad (3-9)$$

Python-ohjelmointikielessä diskreetti Fourier-muunnos lasketaan seuraavalla funktio-kutsulla:

```
numpy.fft.fft(x, n=None, axis=-1, norm=None)
```

missä  $x$  on aikasarja tai signaali, josta diskreetti Fourier-muunnos lasketaan käyttämällä Fast Fourier Transform-algoritmia. Aikasarja tai signaali voi olla myös kompleksinen.

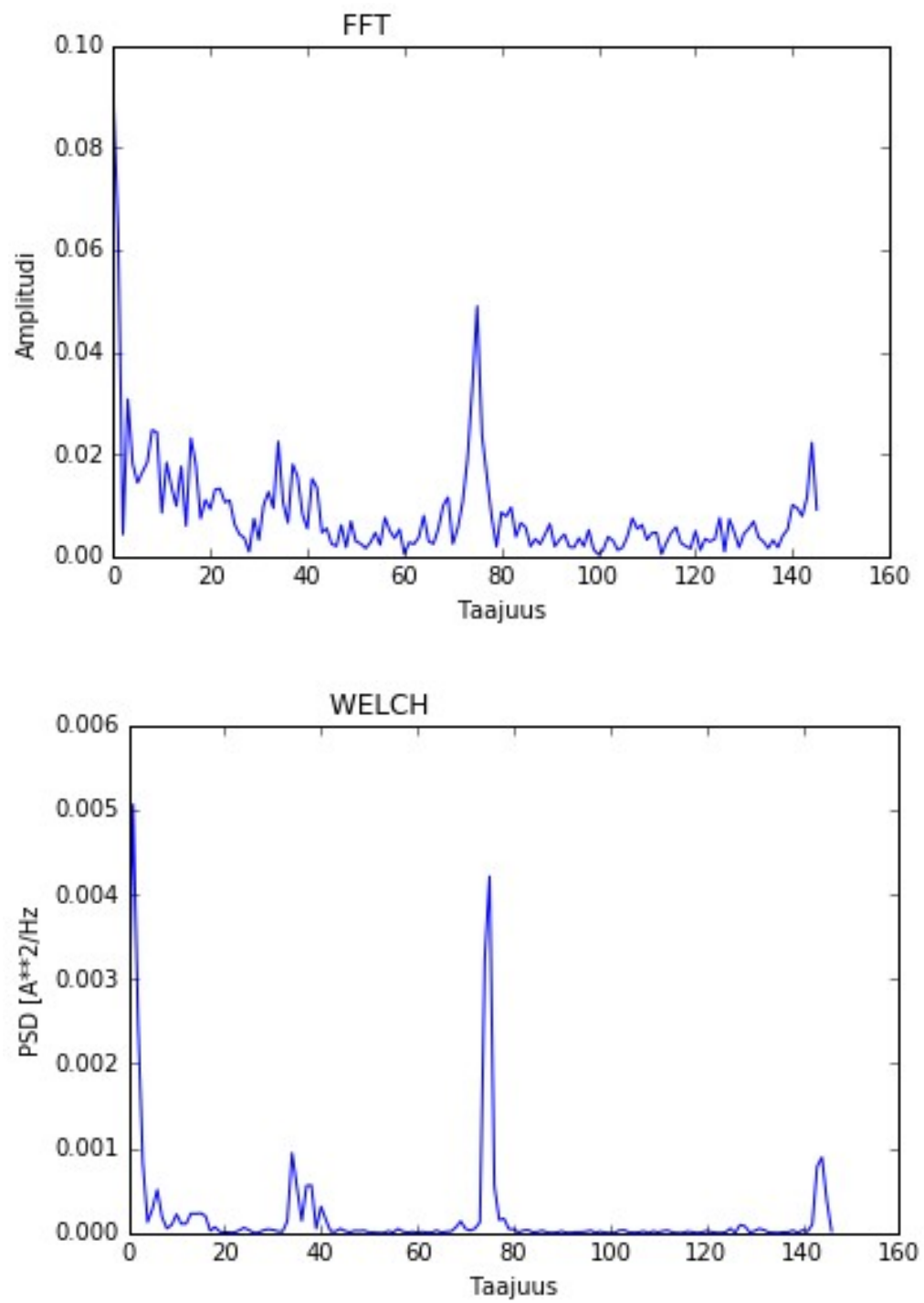
### 3.10 Tehotiheysspektri, Welch'in menetelmä

Welch'in menetelmällä saadaan laskettua signaalista tai aikasarjasta tehotiheysspektri. Tehotiheysspektristä nähdään miten signaalin tai aikasarjan teho on jakautunut taajuu-den mukaan. Welch'in menetelmä on paranneltu menetelmä tehotiheysspektristä ja sen avulla saadaan vähennettyä signaalista tai aikasarjasta aiheutuvaa kohinaa. Welch'in menetelmässä signaali segmentoidaan aikatasossa ja segmenteille tehdään diskreetti Fourier-muunnos. Periodogrammeista lasketaan keskiarvo, mikä vähentää tehotiheys-spektrin arvojen varianssia. [12]

Python-ohjelmointikielessä Welch'in menetelmällä laskettava tehotiheysspektri voidaan suorittaa seuraavalla funktiokutsulla:

```
scipy.signal.welch(x, fs=1.0, window='hann', nperseg=None, noverlap=None, nfft=None, detrend='constant', return_onesided=True, scaling='density', axis=-1),
```

missä parametrilla *nperseg* voidaan määrittää segmentin koko ja parametrilla *noverlap* määritetään montako arvoa menee päällekkäin segmenttien välillä. Kuvassa 6 on Fourier- ja Welch'in menetelmillä saadut spektrit samasta aikasarjasta.



**Kuva 6.** FFT:llä ja Welch'in menetelmällä lasketut taajuusspektrit.

## **4 VIRTAMITTAUSTIETOJEN LUOKITTELUN MENETELMIÄ**

Luokittelu kuuluu koneoppimisen ohjatun oppimisen osa-alueeseen. Ohjatussa oppimisessa tiedetään havaintoarvojen luokat ennalta ja niiden perusteella päätellään tuntemattoman havaintoarvon luokka.

Luokittelussa tieto jaetaan ennalta määrättyihin luokkiin. Luokat määritellään yleensä diskreeteillä arvoilla. Luokittelijat ovat yleensä sovelluskohtaisia ja niiden suunnittelu vaatii tietoa ja kokemusta sovellusalueesta.

Opetustiedosta erotellaan piirteitä, joiden mukaan luokittelija opetetaan tekemään päätöksiä. Alla olevissa kappaleissa kerrotaan piirteiden valinnasta ja esitellään neljä eri luokittelijaa.

### **4.1 Piirteiden määrä ja valinta**

Piirteiden lukumäärä pyritään pitämään mahdollisimman pienenä. Tällä tavoin saadaan luokittelijan päätösalueet pysymään mahdollisimman yksinkertaisina. Yksinkertaiset päätösalueet mahdollistavat luokittelijan opettamisen pienehköllä opetustiedolla. Piirteiden valinta perustuu tiedon luonteeseen ja sitä kuvaavien luokkien määrittelyyn.

Piirteiden valinnassa pyritään etsimään sellaiset piirteet, että ne ovat epäherkkiä mittauksesta johtuvalle virheelle tai kohinalle. Piirteiden valinta aloitetaan järjestelemällä tieto eri luokkien mukaiseen järjestykseen. Järjestetystä tiedosta lähdetään hakemaan eri luokkia parhaiten kuvaavia ja erottelevia piirteitä.

Piirteiden valinnassa voidaan käyttää apuna valittujen piirteiden graafista kuvaimista, josta saadaan yleiskuva siitä, miten eri luokkien valitut piirteet jakautuvat keskenään. Yleiskuvasta havaitaan myös se, miten eri luokkien valitut piirteet jakautuvat muiden luokkien mukaan ja onko niissä päällekkäisyyksiä.

Käytännön piirteiden valinta perustuu kokemukseen sovellusalueesta ja sovellusalueen eri luokkien ominaisuuksien tunnistamiseen. Käytännössä piirteitä voivat olla esimerkiksi keskiarvo, minimi- ja maksiarvot.

## 4.2 K-lähimmän naapurin luokittelija

K-lähimmän naapurin luokittelija on yksi yksinkertaisimmista luokittelu-algoritmeista. Algoritmin ideana on muistaa opetustieto ja ennustaa opetustiedon perusteella uuden datan luokka, perustuen sen lähimpiin naapureihin opetustiedossa. Tämä luokittelutapa perustuu siihen olettamukseen, että tiedon piirteet muodostuvat siten, että lähimpänä olevan naapurin tai naapurien luokka on todennäköisimmin myös uuden tiedon luokka. Lähimmän naapurin etäisyyden määrittämiseen voidaan käyttää erilaisia menetelmiä. Yksi menetelmä on määrittää euklidinen etäisyys lähimpään naapuriin. Euklidinen etäisyys lasketaan alla olevan kaavan mukaan [17]:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (4-1)$$

missä  $x_1$  ja  $x_2$  ovat vektoreita ja alaindekseillä merkityt vektorien alkiot kuvaavat yksittäisiä piirteitä.

K-lähimmän naapurin luokittelijan heikkoutena voidaan pitää sitä, että suurella tietomäärällä muistin tarve kasvaa. Tämä johtuu siitä, että opetustieto pitää ladata muistiin ja uutta tietoa verrataan jokaiseen opetustietoon. Muistin määrän lisäksi k-lähimmän naapurin algoritmin vaatima suoritusaika lisääntyy tiedon määrän lisääntyessä. Näitä ongelmia on pyritty vähentämään käyttämällä likiarvotettua etäisyyttä lähimpään naapuriin. Suosituimmat menetelmät ovat *kd-tree*, *ball tree* ja LSH (*locality-sensitive hashing*). [17]

K-lähimmän naapurin luokittelijan luonti ja käyttö on yksinkertaistettuna seuraavanlainen:

- tiedon käsittely ja luokittelu.
- tiedon jakaminen opetus- ja testausjoukkoon.
- opetetaan kNN-luokittelija opetusjoukolla.
- testataan kNN-luokittelijaa eri k:n arvoilla testausjoukolla.
- valitaan paras k:n arvo ja suoritetaan mahdollisuuksien mukaan luokittelijan validointitestejä erillisellä validointijoukolla.

K-lähimmän naapurin luokittelijan hyvinä puolina voidaan pitää seuraavia ominaisuuksia:

- hyvä luokittelutarkkuus.
- epäherkkä poikkeaville arvoille.
- ei tee oletuksia tiedosta tai havaintojoukosta.

K-lähimmän naapurin luokittelijan huonoina puolina voidaan pitää seuraavia ominaisuuksia:

- laskennallisesti raskas.
- vaatii paljon muistia.
- piirteiden määrä pidettävä pienenä.

Pythonissa kNN-luokittelija on käytettävissä sklearn-paketista ja sen kutsu ja parametrit ovat seuraavassa esimerkissä:

```
class sklearn.neighbors.KNeighborsClassifier(n_neighbors=5, weights='uniform',
      algorithm='auto', leaf_size=30, p=2, metric='minkowski',
      metric_params=None, n_jobs=None, **kwargs),
```

missä *n\_neighbors* määrittelee naapureiden lukumäärän, (oletuksena käytetään viittä), *weights* parametrilla valitaan käytettävä painotusfunktio, (oletuksena käytetään 'uniform' funktiota, joka painottaa kaikkia pisteitä samanarvoisesti) ja parametrilla *algorithm* valitaan käytettävä algoritmi lähimmän naapurin laskennassa (oletuksena käytetään 'auto' valintaa, joka yrittää määrittellä parhaan mahdollisen algoritmin sen perusteella, mitä *fit*-funktioon on syötetty).

Pythonissa sklearn-paketissa oleva kNN-luokittelija tukee seuraavia algoritmeja:

- *ball\_tree*.
- *kd\_tree*.
- *brute*, joka käyttää niin sanottua *brute-force* etsintää (käydään kaikki lävitse).

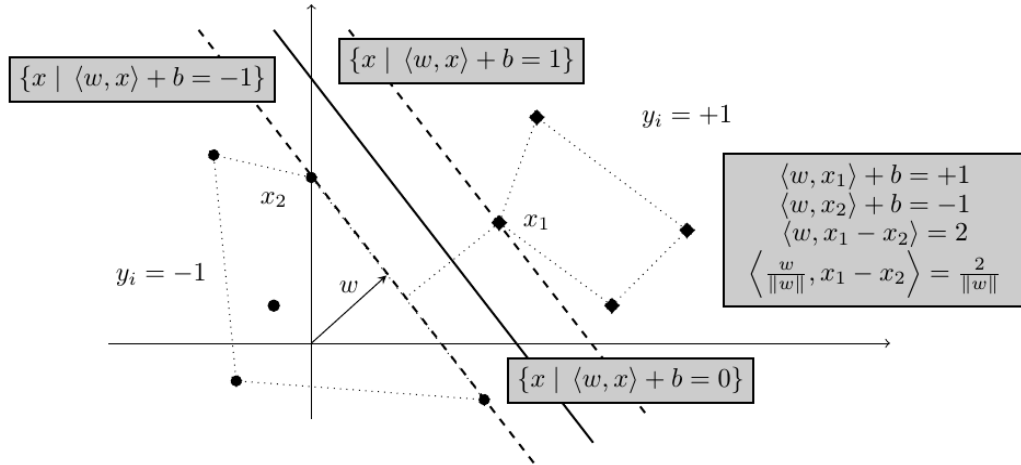
### 4.3 Tukivektori-kone ( Support Vector Machine, SVM)

Tukivektori-kone on tilastollisen koneoppimisen teoriaan perustuva lineaarinen luokittelija. Tukivektori-kone on kahden luokan luokittelija ja oletuksena luokat ovat lineaarisesti separoituvia. Tukivektori-koneen toiminta perustuu tukivektoreihin, jotka ovat kahden havaintojoukon väliin sovitettuja tasoja. Tukivektorit pyritään sovittamaan siten, että yhdensuuntaisten tasojen välimatka on mahdollisimman suuri. Sovittamisessa huolehditaan myös siitä, että yksikään havaintojoukon piste tai näyte ei jää tasojen väliin. Luokat erottavaa tasoa kutsutaan hypertasoksi. Hypertason yhtälö on seuraava [18]:



$$w \cdot x + b = 0 \quad (4-3)$$

missä  $w$  on tason normaalivektori ja  $b$  skalaari kerroin. Kuvassa 7 on havainnollistettu tukivektoreita ja hypertasoa.



**Kuva 7.** Tukivektorit ja hypertaso [14]

Havaintojoukot jotka eivät ole lineaarisesti separoituvia muodostavat ongelman käytettäessä lineaarisia tasoja. Epälineaarisesti separoituvan havaintojoukon kanssa käytetään joustavan marginaalin toteutusta. Joustava marginaali toteutetaan niin sanottujen *slack*-vakioiden avulla, jotka ilmaisevat havaintojoukon pisteen etäisyyden siitä päätöspinnasta johon se kuuluisi. Joustavan marginaalin luokittelussa käytetään erilaisiin ydinfunktioihin perustuvia ratkaisuja. [18]

Tukivektorikoneessa on ydinfunktioita (*kernel*), joita käytetään epälineaarisesti separoituvan havaintojoukon luokitteluun. Ydinfunktioilla havaintojoukko jaotellaan moniulotteisissa tasoissa. Tällä tavoin saadaan epälineaarisesti separoituva havaintojoukko jaoteltua lineaarisesti moniulotteisessa avaruudessa. Yleisimmät ydinfunktiot ovat *linear*, *polynomial* ja RBF (*radial basis function*) [18].

Tukivektorikoneen hyvinä puolina voidaan pitää seuraavia ominaisuuksia [15]:

- ei ylisovitu helposti.
- laskennallisesti tehokas.
- tulosten tulkinta helppoa.

Tukivektorikoneen huonoina puolina voidaan pitää seuraavia ominaisuuksia [15]:

- herkkä parametrien valinnalle ja ydinfunktion valinnalle (automaattiset valinnat voivat ylisovittua).
- alun perin käsittelee vain binääristä luokittelua.
- mallia vaikeahko tulkita.
- vain numeeriselle tiedolle tai havaintojoukolle.

Python-ohjelmointikielessä tukivektorikone-luokittelija on käytettävissä sklearn-paketista ja sen kutsu ja parametrit ovat seuraavassa esimerkissä:

```
class sklearn.svm.SVC(C=1.0, kernel='rbf', degree=3, gamma='auto_deprecated',
coef0=0.0, shrinking=True, probability=False, tol=0.001, cache_size=200,
class_weight=None, verbose=False, max_iter=-1, decision_function_shape='ovr',
random_state=None)
```

missä *kernel*-parametrilla voidaan määritellä käytettävä ydinfunktio, parametrilla *degree* asetetaan *polynomial*-ydinfunktiota käytettäessä *polynomial*-ydinfunktion aste ja parametrilla *gamma* asetetaan ydinfunktion kerroin käytettäessä ydinfunktiota RBF, *polynomial* tai *sigmoid*.

Python-ohjelmointikielessä sklearn-paketissa oleva tukivektorikone tukee seuraavia ydinfunktiota:

- linear  $\langle x, x' \rangle$ .
- polynomial  $(\gamma \langle x, x' \rangle + r)^d$ , missä  $d$  on parametrin aste ja  $r$  on parametrin *coef0* arvo.
- rbf  $\exp(-\gamma \|x - x'\|^2)$ , missä  $\gamma$  on parametrin *gamma* arvo.
- sigmoid  $(\tanh(\gamma \langle x, x' \rangle + r))$ , missä  $r$  on parametrin *coef0* arvo.

Näiden lisäksi voidaan käyttää omia ydinfunktioita. Oman ydinfunktion käyttö määritellään *kernel*-parametrilla alustettaessa tukivektorikone.

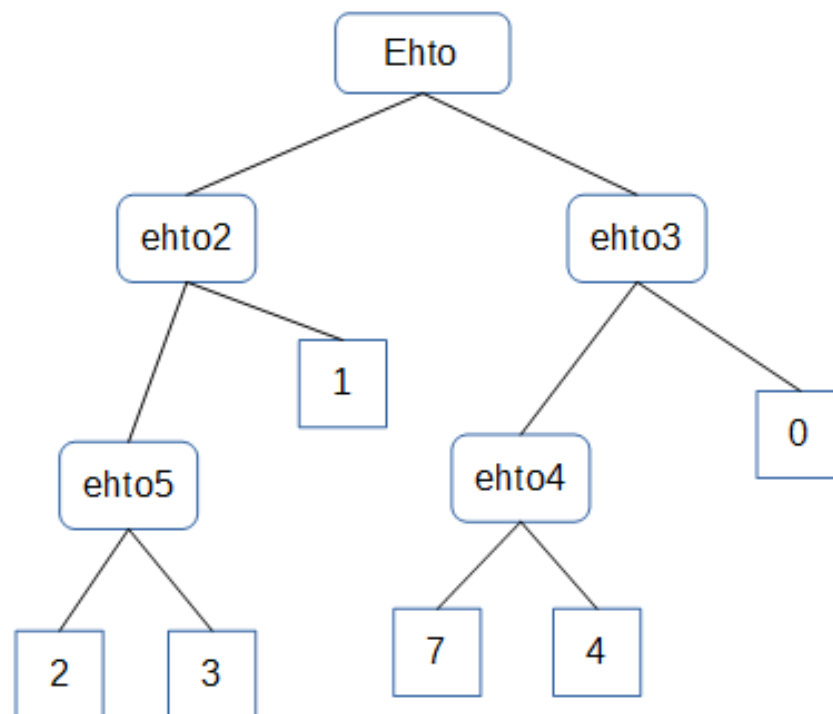
## 4.4 Päättöspuu-luokittelija

Päättöspuu-luokittelija on puurakenteinen luokittelija, jonka oksat kuvaavat testiä tai ehtoa ja lehtisolmut kuvaavat kategoriaa tai luokkaa. Kategoriat ja luokat ovat päättöspuu-luokittelijan lopputuloksia ja oksat ovat testejä tai ehtoja, jotka ohjaavat päätöksen tekoa

puurakenteessa. Päättöspuulla voidaan luokitella monenlaista tietoa, joka voidaan jakaa eri luokkiin. Päättöspuun tyypillisiä ominaisuuksia ovat [15]:

- testi tai ehto voi olla usean muuttujan tai yhden muuttujan testi tai ehto.
- testillä tai ehdolla voi olla kaksi tai useampia lopputuloksia, jos kaikilla testeillä tai ehdoilla on aina kaksi lopputulosta on kyseessä binäärinen päättöspuu.
- kategoriat tai luokat voivat olla numeerisia tai kuvaavia.
- kategorioita ja luokkia voi olla useita.

Kuvassa 8 on yksinkertaistettu esimerkki päättöspuu-luokittelijasta.



**Kuva 8.** Päättöspuu-luokittelijan yksinkertaistettu esimerkki.

Päättöspuiden rakentamiseen on kehitetty useita algoritmeja ja yleisimpiä niistä ovat ID3, C4.5 ja CART. C4.5 algoritmista on saatavilla paranneltu versio C5.0 ja sen etuina ovat pienempi muistin käyttö, pienemmät päätössäännöt ja parempi tarkkuus. Algoritmit huolehtivat päättöspuun rakentamisesta ja sen karsimisesta. Päättöspuu-luokittelijat käyttävät yleisesti tiedon jakamisen kriteerinä kahta eri menetelmää, entropian kasvua informaatiossa tai epäpuhtauden kasvua informaatiossa. [18]

Entropian kasvu informaatiossa lasketaan kaavalla (*cross-entropy*):

$$H(p, q) = - \sum_x p(x) \log q(x) \quad (4-6)$$

missä  $p$  ja  $q$  ovat havaintojoukon tai tiedon todennäköisyysjakaumat.

Epäpuhtauden kasvu  $L$ -luokkaiselle havaintojoukolle lasketaan kaavalla (*Gini*):

$$G(p) = 1 - \sum_{l=1}^L p_l^2 \quad (4-7)$$

missä  $l \in (1, 2, \dots, L)$  ja  $p_l$  on luokan  $l$  osuus kaikista arvoista.

Luokitteluvirhe lasketaan kaavalla:

$$E(p) = 1 - \max(p_l) \quad (4-8)$$

missä  $p_l$  on luokan  $l$  osuus luokasta  $L$

Päätöspuiden hyvinä puolina voidaan pitää seuraavia ominaisuuksia:

- helpohko ymmärtää ja tulkita, koska päätöspuu voidaan visualisoida.
- ei vaadi paljoa tiedon tai havaintojoukon esikäsittelyä.
- voidaan käsitellä numeerista tai kategorista tietoa.

Päätöspuiden huonoina puolina voidaan pitää seuraavia ominaisuuksia:

- ylisovittuu helposti.
- tiedon tai havaintojoukon piirteistä johtuen päätöspuu voi vääristyä jonkin luokan mukaan (vaatii opetustiedon tarkastelua).
- optimaalisen päätöspuun rakentaminen on niin sanottu *np*-kova ongelma.

Python-ohjelmointikielessä päätöspuu luokittelija on käytettävissä *sklearn*-paketista ja sen kutsu ja parametrit ovat seuraavassa esimerkissä:

```
class sklearn.tree.DecisionTreeClassifier(criterion='gini', splitter='best',
max_depth=None, min_samples_split=2, min_samples_leaf=1,
```

```
min_weight_fraction_leaf=0.0,      max_features=None,      random_state=None,  
max_leaf_nodes=None,      min_impurity_decrease=0.0,      min_impurity_split=None,  
class_weight=None, presort=False)
```

missä parametrilla *criterion* määritellään käytettävä tiedon jakamisen kriteeri (oletuksena käytetään *gini*-arvoa, mikä jakaa tiedon epäpuhtauden kasvun mukaan). Parametrilla *max\_depth* määritellään puun syvyys, eli miten paljon oksia puuhun muodostuu (oletuksena käytetään *none*, joka laajentaa puuta kunnes lehdet ovat niin sanotusti puhtaita tai *min\_samples\_split*-parametrin arvo saavutetaan jakoehdossa).

## 4.5 Lineaarinen diskriminanttianalyysi (LDA)

Lineaarinen diskriminanttianalyysi on lineaarinen luokittelija. Luokittelu perustuu lineaarikombinaatioihin, joiden avulla tieto tai havaintojoukko jaetaan lineaarisesti eri luokkiin. Lineaarinen diskriminanttianalyysi olettaa, että luokkien ehdolliset todennäköisyysjakaumat ovat normaalijakautuneita. Lisäksi eri luokkien todennäköisyysjakauksilla pitäisi olla sama kovarianssimatriisi, sekä piirteet oletetaan tilastollisesti riippumattomiksi. Käytännössä nämä kaikki ehdot toteutuvat harvoin, mutta lineaarinen diskriminanttianalyysi voi silti suoriutua riittävän hyvin. Lineaarinen diskriminanttianalyysi antaa parhaat tulokset standardoidulla tiedolla tai havaintojoukolla. [18]

Lineaarisen diskriminanttianalyysin hyvinä puolina voidaan pitää seuraavia ominaisuuksia:

- yksinkertainen luokittelija (lähimmän luokan keskipiste).
- lineaariset päätösrajat.

Lineaarisen diskriminanttianalyysin huonoina puolina voidaan pitää seuraavia ominaisuuksia:

- lineaariset rajat eivät ole aina riittäviä.
- tiedon tai havaintojoukon vaatimukset jakauman ja kovarianssin suhteen.

Python-ohjelmointikielessä lineaariseen diskriminanttianalyysiin perustuva luokittelija on käytettävissä *sklearn*-paketista ja sen kutsu ja parametrit ovat seuraavia:

```
class sklearn.discriminant_analysis.LinearDiscriminantAnalysis(solver='svd',
shrinkage=None, priors=None, n_components=None, store_covariance=False,
tol=0.0001)
```

missä parametrilla *solver* määritellään käytettävä ratkaisijan tyyppi (oletuksena käytetään *svd*:tä). Parametrilla *shrinkage* määritellään käytettävä kutistaja (oletuksena kutistajaa ei käytetä).

Python-ohjelmointikielessä sklearn-paketissa oleva lineaarinen diskriminanttianalyysi-luokittelija tukee seuraavia ratkaisijoita:

- *svd*, pääakselihajotelma ( *Singular value decomposition*). Tämä ratkaisija ei laske kovarianssimatriiseja ja on siksi suositeltava tiedolle tai havaintojoukolle joka sisältää runsaasti piirteitä.
- *lsqr*, pienimmän neliösumman menetelmä ( *Least squares solution*). Tätä ratkaisijaa voidaan käyttää kutistajan kanssa.
- *eigen*, eigenvektoreiden hajotelma ( *Eigenvalue decomposition*). Tätä ratkaisijaa voidaan käyttää kutistajan kanssa.

Python-ohjelmointikielessä sklearn-paketissa oleva diskriminanttianalyysi-luokittelija tukee seuraavia kutistajia:

- *auto*, käyttää Ledoit-Wolf lemmaa kutistamiseen.
- liukuluku 0 ja yhden välillä, jota käytetään kutistajan arvona.

Python-ohjelmointikielen sklearn-paketissa oleva lineaarinen diskriminanttianalyysi luokittelijan kutistajan avulla voidaan parantaa kovarianssimatriisien ennustettavuutta, kun tiedon tai havaintojoukon opetustiedon määrä on pieni verrattuna piirteiden määrään.

#### 4.5.1 Neliöllinen diskriminanttianalyysi (QDA)

Neliöllinen diskriminanttianalyysi ei oleta, että kovarianssit ovat samoja jokaisella luokalla. Päätosrajapinnat eivät ole lineaarisia ja ne muodostetaan toisen asteen polynomien avulla. [18]

Python-ohjelmointikielessä neliölliseen diskriminanttianalyysiin perustuva luokittelija on käytettävissä sklearn-paketista ja sen kutsu ja parametrit ovat seuraavassa esimerkissä:

```
class sklearn.discriminant_analysis.QuadraticDiscriminantAnalysis  
(priors=None, reg_param=0.0, store_covariance=False, tol=0.0001, store_covarian-  
ces=None)
```

missä parametrilla *priors* voidaan asettaa luokkien oletuksia ja parametrilla *reg\_param* voidaan säätää kovarianssimatriisien estimaattia.

## 5 VIRTAMITTAUSTIETOJEN ANALYSOINNIN TULOKSET

Tässä analyysissä tutkittiin yhteensä 116 konetta jokaisen kolmen pääakselin osalta. Käytännön työssä kerättiin, koostettiin ja jalostettiin virtamittaustiedot sellaiseen muotoon, että niitä voitiin käsitellä ohjelmallisesti. Koostamisen ja jalostamisen jälkeen virtamittaustiedoista määritettiin piirteitä, joiden mukaan tehtiin luokittelua eri levytyökonemallien mukaan. Eri levytyökonemalleista luokiteltiin erikseen jokainen pääakseli. Alla olevissa kappaleissa on kerrottuna ja kuvattuna tarkemmin eri osa-alueiden työmenetelmiä sekä saatuja tuloksia.

### 5.1 Levytyökoneiden virtamittausten koostaminen

Ennen tietojen koostamista tehtiin valinta levytyökonemalleista, joista tietoja kerätään ja koostetaan. Valituista levytyökonemalleista tehtiin vielä jaottelu eri sarjojen mukaan. Valitut levytyökonemallit olivat lävistävä, lineaari ja yhdistelmä-levytyökone. Nämä mallit olivat kattava kokonaisuus valmistettavista levytyökonemalleista ja niiden virtamittaustietojen analysoinnilla saatiin kattava otanta eri levytyökonemallien osalta.

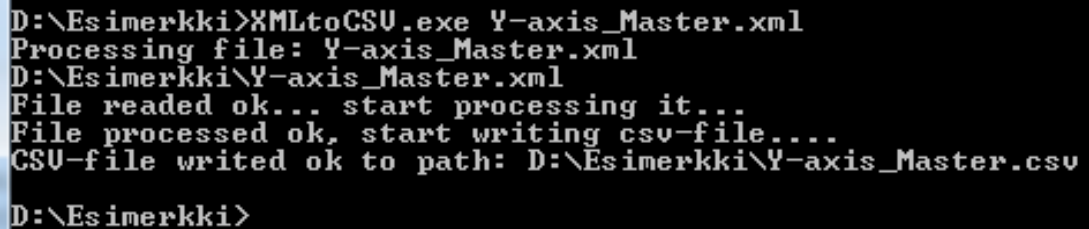
Pilvipalvelusta ladattiin eri konemallien virtamittaustiedot tuotannon valmistuksen päätteeksi tehdyistä virtamittauksista. Ladatut virtamittaustiedot olivat xml formaatissa tallennettuna. Xml-formaatissa olevat tiedot päätettiin muuntaa csv-formaatin mukaiseksi, koska se on helpommin siirrettävä ja käsiteltävä formaatti eri ohjelmistojen välillä.

Virtamittaustietojen tallennusformaattiin oli tehty parannuksia ja nämä parannukset olivat muuttaneet virtamittausten tallentamiseen käytetyn xml-formaatin rakennetta. Pilvipalvelusta haetuista virtamittausarvoista ei pystytty etukäteen määrittelemään missä formaatissa xml-tiedostot olivat. Tämä aiheutti sen, että päätettiin tehdä erillinen ohjelmisto, joka osaa ottaa huomioon molemmilla formaateilla talletetut xml-tiedostot.

Pilvipalvelusta haettujen virtamittaustietojen muuntaminen xml-formaatista csv-formaattiin tehtiin c#-ohjelmointikielellä. Virtamittausten formaatin muunnokseen xml-formaatista csv-formaattiin tehtiin yksinkertainen komentokehoteesta suoritettava ohjelma. Kehitetty ohjelmisto tuotti jokaisesta pääakselin xml-tiedostosta csv-tiedoston. Csv-tiedostosta jätettiin pois kaikki tarpeettomat lisätiedot, joita oli alkuperäisissä xml-tiedostoissa, koska niitä ei tarvittu virtamittaustietojen analysoinnissa. Kuvassa 9 on esi-



merkki onnistuneesta pääakselin xml-formaatissa olevan virtamittauksen muunnoksesta csv-formaattiin.



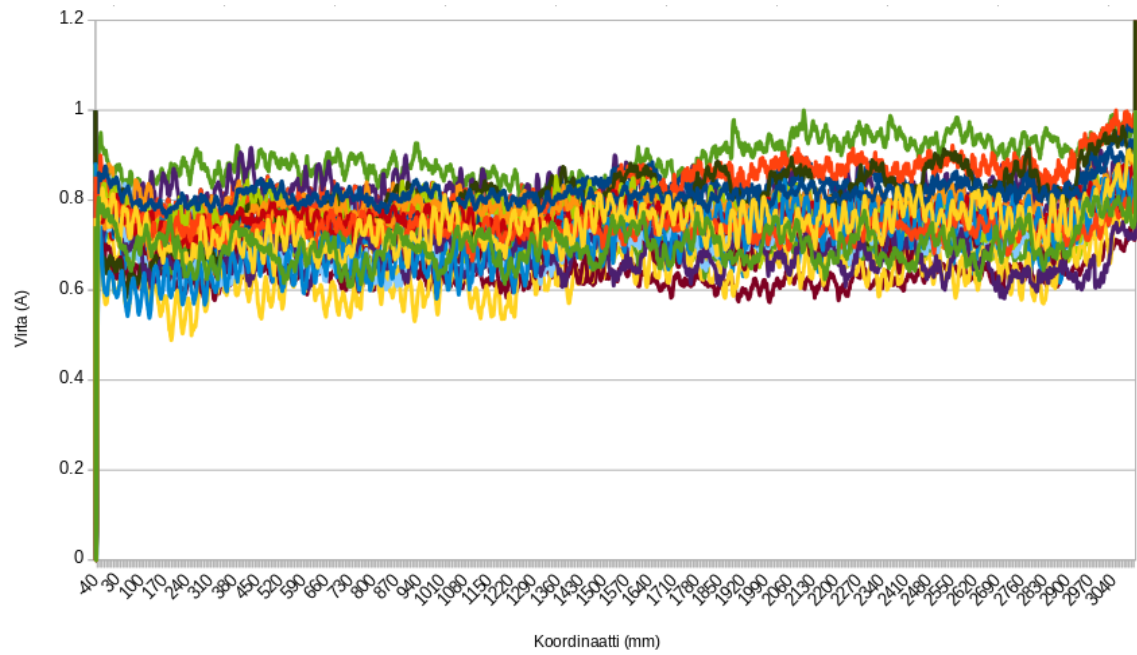
```
D:\Esimerkki>XMLtoCSV.exe Y-axis_Master.xml
Processing file: Y-axis_Master.xml
D:\Esimerkki\Y-axis_Master.xml
File readed ok... start processing it...
File processed ok, start writing csv-file....
CSV-file writed ok to path: D:\Esimerkki\Y-axis_Master.csv
D:\Esimerkki>
```

*Kuva 9. Xml-formaatin muuntaminen csv-formaattiin.*

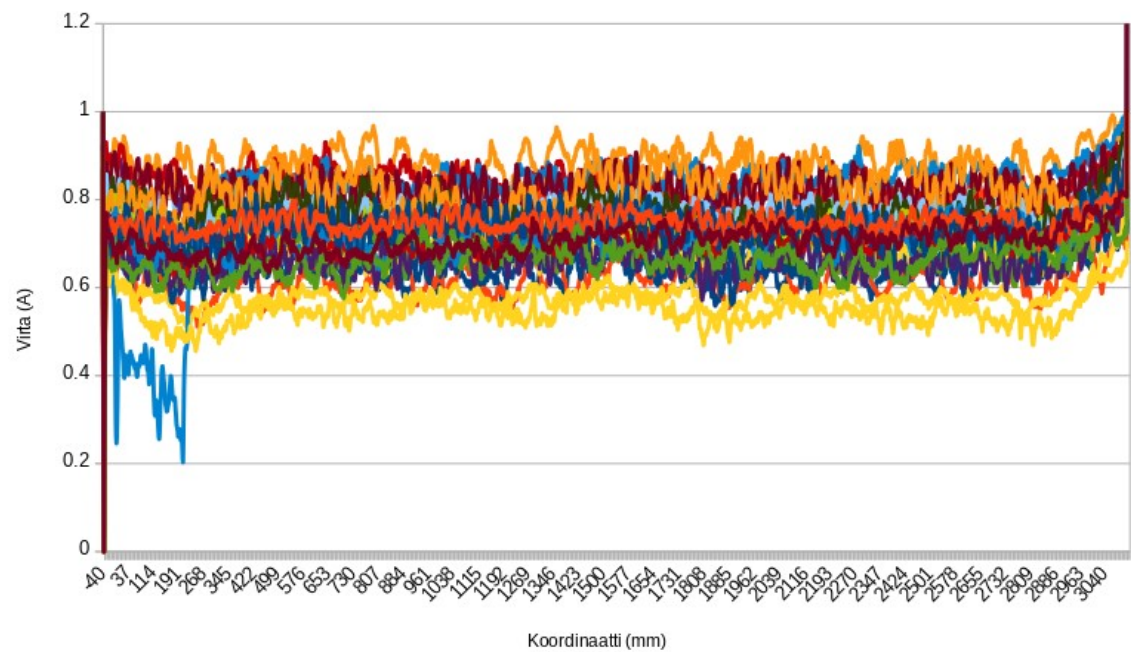
Jokaisen konemallin osalta koostettiin pääakseleittain erilliset csv-tiedostot. Näistä erillisistä csv-tiedostoista koostettiin yhtenäinen csv-tiedosto, jossa on kaikki saman konemallin tiedot pääakseleittain. Näin saatiin analyysia varten eriteltyä jokainen konemalli pääakseleiden mukaan. Koostamisen tuloksena saatiin jokaisesta eri konemallista pääakseleiden tiedot sisältävät csv-tiedostot, joita voitiin käsitellä eri ohjelmistoilla.

Koostamisen jälkeen saadut tiedot visualisointiin graafisesti ja perehdyttiin tarkemmin eri akselien virtamittausten kuvaajiin. Kuvaajien perusteella kerättiin asiantuntija kommentteja pääakseleiden virtamittausten kuvaajista. Kommenttien perusteella saatiin aineistosta poistettua virhemittaukset. Virhemittauksia olivat muun muassa liian lyhyet virtamittaussarjat ja epätäydellistä tietoa sisältäneet virtamittaukset. Kesken jääneet mittaukset ja epäonnistuneet mittaukset pystyttiin tunnistamaan konemallin teknisten ominaisuuksien perusteella.

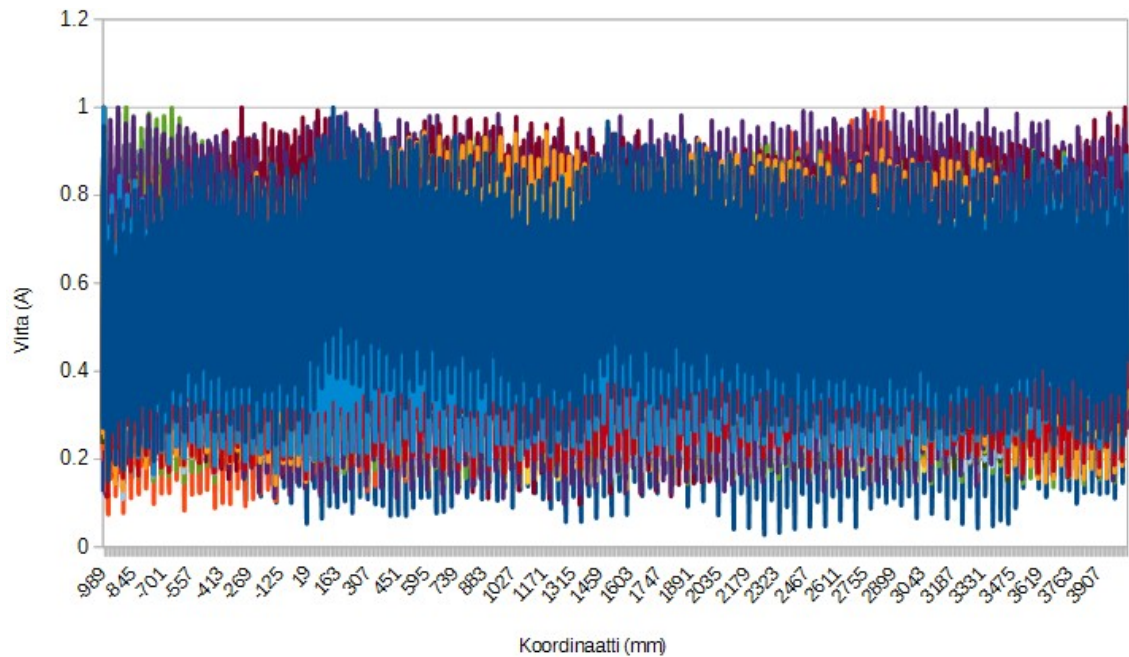
Kuvassa 10 on lävistävän konemallin yhden pääakselin kuvaajia 28 kappaletta, kuvassa 11 on yhdistelmä-konemallin yhden pääakselin kuvaajia 29 kappaletta ja kuvassa 12 on lineaarisen-konemallin yhden pääakselin kuvaajia 25 kappaletta.



**Kuva 10.** Lävistävän levytyökoneen yhden pääakselin virtakuvaajia.



**Kuva 11.** Yhdistelmä-levytyökoneen yhden pääakselin kuvaajia.



**Kuva 12.** Lineaari-levytyökoneen yhden pääakselin kuvaajia.

Lävistävän ja yhdistelmä-levytyökoneen kuvissa 10 ja 11 on pääakseleilla sama mekaaninen rakenne ja niitä ohjataan samankokoisella servomootorilla. Kuvissa 10 ja 11 on lisäksi sama pääakselin mitoitus, joten virtamittausarvojen tulisi olla yhteneväiset. Kuvista 10 ja 11 nähdään, että virtamittausarvot ovat yhteneväiset, mutta yhdessä yhdistelmä-levytyökoneen virtamittausarvossa on selvä poikkeama pääakselin alkupäässä.

Lineaari-levytyökoneen virtamittausarvojen ero tulee pääakselin ohjaukseen käytettävän lineaarimoottorin ominaisuuksista. Lineaarimoottori poikkeaa rakenteeltaan ja ohjaukseltaan servomoottorista. Tästä erosta johtuen lineaarisen-levytyökoneen virtamittauskäyrät poikkeavat huomattavasti lävistävän levytyökoneen ja yhdistelmä-levytyökoneen pääakseleiden virtamittausarvojen muodosta.

## 5.2 Raja-arvojen määrittely

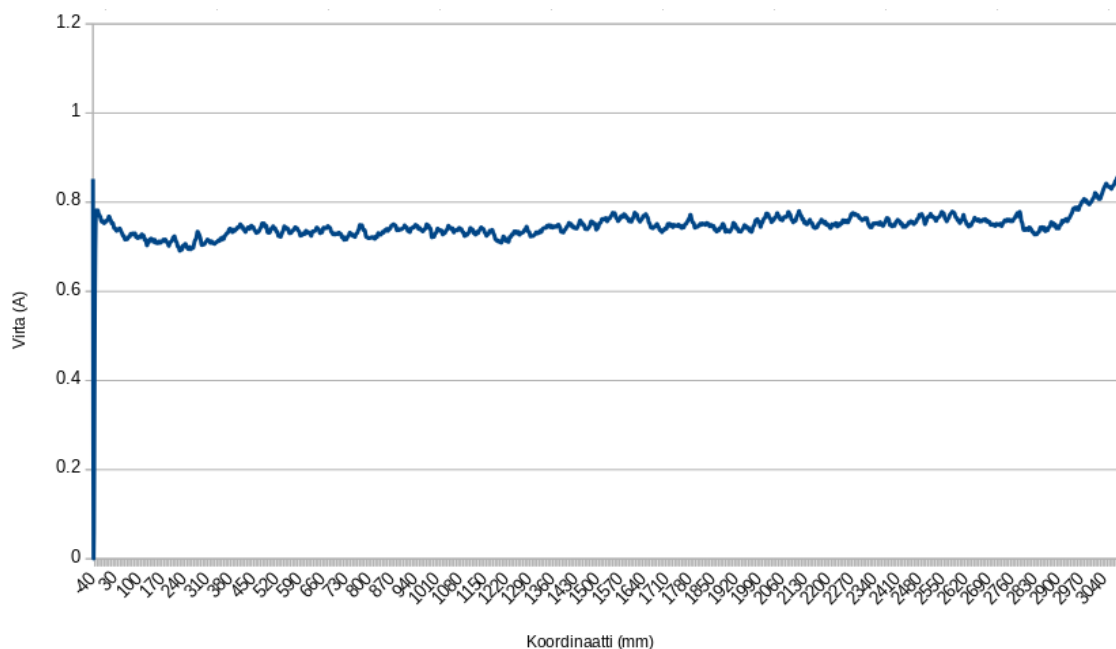
Koostamisen jälkeen levytyökoneiden virtamittaukset olivat käyttökelpoista tietoa, josta oli poistettu kaikki mittausprosessissa tapahtuneet virheelliset virtamittaukset. Asiantuntija kommenttien perusteella tehtiin virtamittauskuvaajien jako kolmeen eri ryhmään. Näistä kolmesta ryhmästä käytetään tässä analyysissä nimityksiä 'ok', 'tarkista' ja 'ongelma'. 'Ok'-ryhmään kuuluivat kaikki asiantuntijoiden mielestä hyvät virtamittausku-

vaajat. 'Tarkista'-ryhmään kuuluivat kaikki vähän poikkeavat virtamittauskuvaajat, mutta eivät ilmiselvät 'ongelma'-ryhmään kuuluvat virtamittauskuvaajat. 'Ongelma'-ryhmään kuuluivat kaikki virtamittauskuvaajat, jossa oli havaittavissa ilmiselvä ongelma.

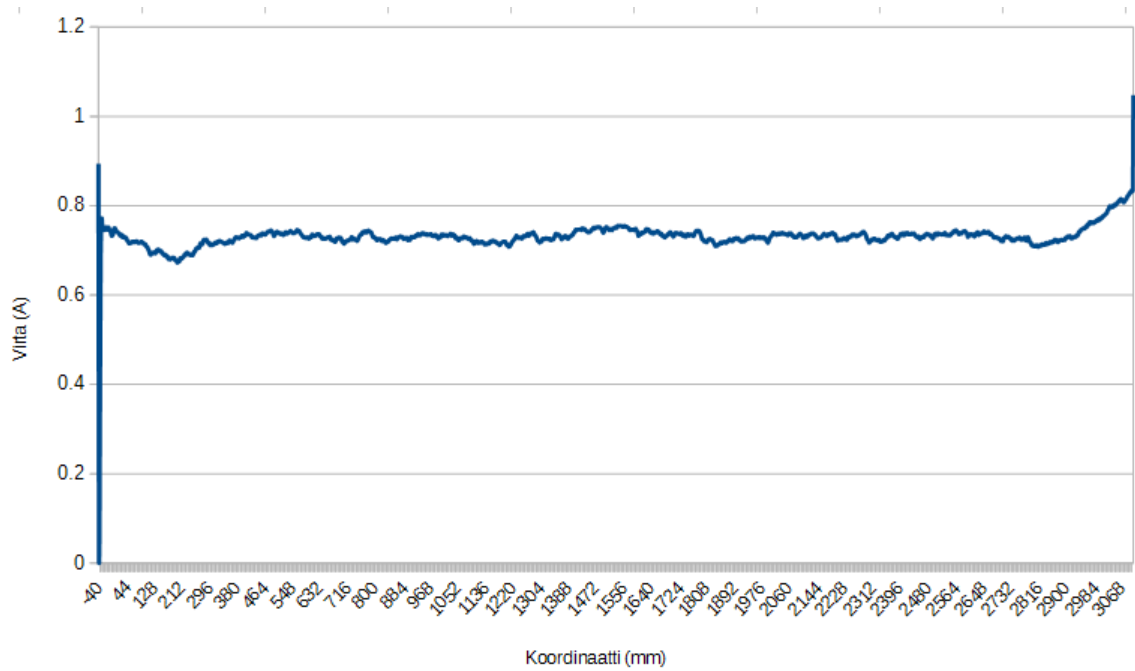
Jokainen levytyökonemalli jaettiin pääakseleiden mukaisiin ryhmiin ja nämä ryhmät jaettiin kolmeen virtamittauskuvaajan muotoa kuvaaviin ryhmiin. Tällä tavoin saatiin jokainen pääakseli ryhmiteltyä joko 'ok'-, 'tarkista'- tai 'ongelma'-ryhmään. Nämä ryhmäjaot toimivat myös luokittelun perustana. Raja-arvojen määrittelyssä tärkein ryhmä on 'ok'-ryhmä, jolla saadaan määriteltyä jokaisen levytyökonemallin hyväksyttävät raja-arvot pääakseleittain. Analyysissa päädyttiin muutamien kokeilujen perusteella laskemaan keskiarvokuvaaja jokaisen 'ok'-ryhmän pääakseleiden osalta. Tällä tavoin saatiin parhaiten kuvattua virtamittauskuvaajan muoto ja trendi pääakselikohtaisesti.

Keskiarvotettua 'ok'-ryhmän virtamittauskuvaajaa käytettiin vertailuarvona, kun 'tarkista'- ja 'ongelma'-ryhmiä arvioitiin uudelleen. Tällä tavoin saatiin paremmin esille virtamittauskuvaajien eroavaisuuksia.

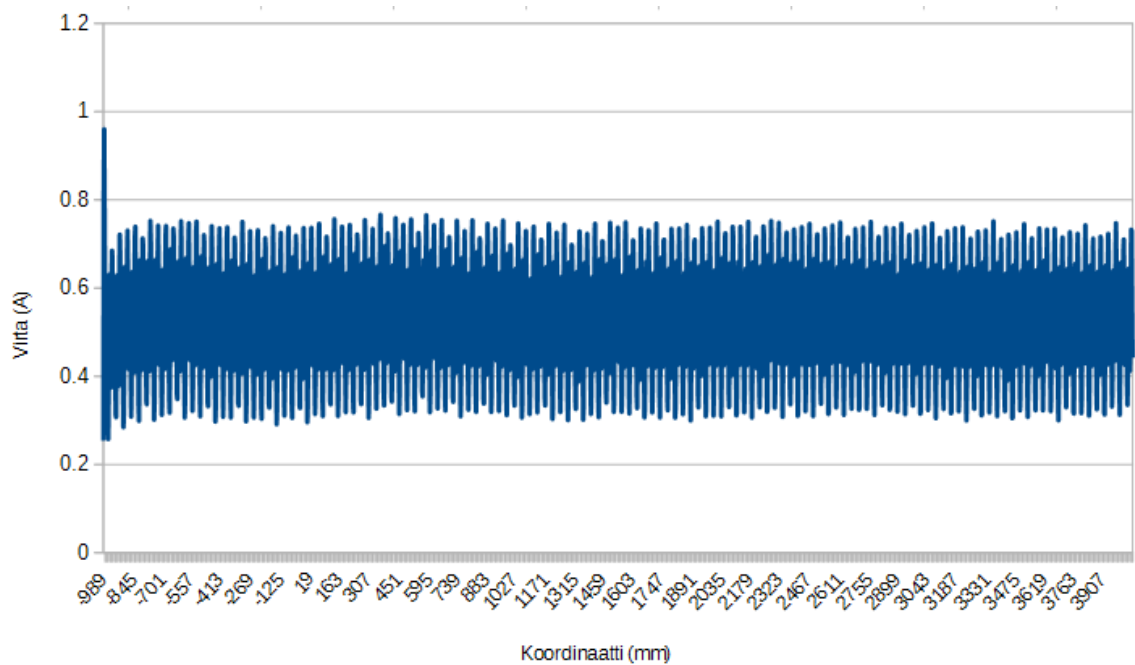
Kuvassa 13 on lävistävän levytyökoneen yhden pääakselin keskiarvotettu 'ok'-ryhmän virtamittauskuvaaja, kuvassa 14 on yhdistelmä-levytyökoneen yhden pääakselin keskiarvotetun 'ok'-ryhmän virtamittauskuvaaja ja kuvassa 15 on lineaari-levytyökoneen yhden pääakselin keskiarvotettu 'ok'-ryhmän virtamittauskuvaaja.



**Kuva 13.** Lävistävän levytyökoneen yhden pääakselin 'ok'-ryhmän keskiarvotettu virtamittauskuvaaja.



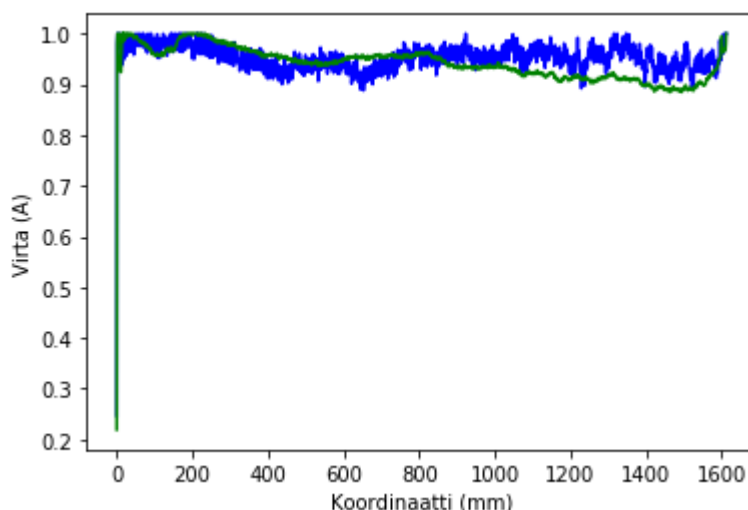
**Kuva 14.** Yhdistelmä-levytyökoneen yhden pääakselin 'ok'-ryhmän keskiarvotettu virtamittauskuvaaja.



**Kuva 15.** Lineaari-levytyökoneen yhden pääakselin 'ok'-ryhmän keskiarvotettu virtamittauskuvaaja.

Lävistävän levytyökoneen ja yhdistelmä-levytyökoneen 'ok'-ryhmän keskiarvotetut virtamittauskäyrät ovat yhteneväiset, koska molempien levytyökoneiden pääakseleiden mekaaninen ja sähköinen rakenne on samanlainen yhden pääakselin osalta. Lineaari-levytyökoneen virtamittauksen eroavaisuus tulee lineaarimoottorin ominaisuuksista ja sen erilaisesta ohjauksesta.

Pääakseleiden 'ok'-ryhmän kuvaajista määriteltiin jokaiselle levytyökonetyypille raja-arvot pääakseleittain. 'Ok'-ryhmän virtamittauskuvaajia käytettiin myös luokittelun apuna määritettäessä vertailuarvoja eri luokkien kesken. Kuvassa 16 on esimerkki, missä 'ok'-luokan keskiarvotettua kuvaajaa käytetään vertailuarvona yksittäiselle pääakselin virtamittauskuvaajalle.



**Kuva 16.** Yksittäisen virtamittauskuvaajan vertailu keskiarvotetun kuvaajan kanssa.

Kuvassa 16 vihreällä värillä on keskiarvotettu vertailukuvaaja ja sinisellä värillä yksittäisen pääakselin virtamittauskuvaaja. Vertailukuvaajasta saadaan hyvin selville, miten virta-arvot ovat jakautuneet pääakselin pituuden mukaan ja miten yksittäisen pääakselin virtamittauskäyrän muoto noudattaa vertailukuvaajan muotoa.

Raja-arvojen määrittelyssä ilmeni, että yhden pääakselin virtamittausprosessissa on ulkoisen muuttujan aiheuttama vääristymä. Ulkoinen muuttuja pääsee vaikuttamaan suoraan virtamittauksen arvoihin ja näin ollen aiheuttaa ennalta määrittelemättömän vääristymän virtamittauksen tulokseen. Ulkoinen muuttuja säätaa pääakseliin vaikuttavaa voimaa, mikä vastustaa pääakselin liikettä virtamittauksen aikana. Virtamittauksista ilmeni, että jokaisella levytyökoneella on tässä säädössä eroavaisuuksia. Ulkoisen muuttujan säädön ja asetusarvon muutoksia testattiin erikseen mittaamalla virtamittauksia eri

säätöarvoilla. Näistä virtamittauksista saatiin selville, että pienehköt erot ulkoisen muuttujan säädössä aiheuttivat isomman muutoksen virtamittauksen arvoihin. Ulkoisesta muuttujasta johtuen kaikista levytyökonetyypeistä ei voida määritellä yhden pääakselin raja-arvoja. Raja-arvon määrittäminen kyseessä olevalle pääakselille vaatii virtamittausprosessin muutosta. Muutosta ei saatu tehtyä tämän diplomityön aikana, koska muutos kohdistuu levytyökoneen mekaaniseen rakenteeseen.

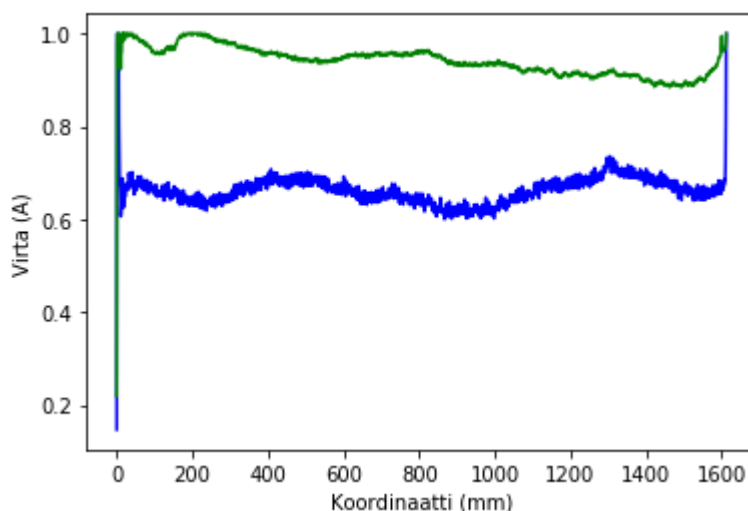
### 5.3 Piirteiden valinta

Raja-arvojen määrittelyn jälkeen lähdettiin kartoittamaan virtamittausten piirteitä. Ensimmäisenä tarkasteltiin virtakäyrien amplitudin muotoa ja koostettiin kaikkien kolmen eri luokan virtakäyrien amplitudien muodot yhteen vertailua varten. Asiantuntijoiden avustuksella arvioitiin virtakäyrien amplitudien muodoista niitä piirteitä ja ominaisuuksia mitkä kuvaisivat parhaiten eri luokkien eroavaisuuksia toisiinsa nähden.

Ensimmäiseksi lähdettiin tutkimaan pääakselien virtakäyrien keskiarvoista virta-arvoa. Keskiarvoinen virta-arvo ilmaisee sen, miten paljon voimaa tarvitaan aikaansaamaan tasaisella nopeudella tehtävä pääakselin virtamittaus. Levytyökonetyypeittäin tutkittiin jokainen pääakseli erikseen. 'Tarkista'- ja 'ongelma'-ryhmän keskiarvotettua virta-arvoa verrattiin 'ok'-ryhmän keskiarvotettuun vertailuarvoon. 'Tarkista'- ja 'ongelma'-ryhmän virtakäyrän amplitudin keskiarvotettu virta-arvo ei ollut riittävän hyvä erottelemaan 'tarkista'- ja 'ongelma'-ryhmän tasoeroa 'ok'-ryhmän keskiarvoisen virta-arvon tasosta.

'Tarkista'- ja 'ongelma'-ryhmään kuuluvien virtamittausten tasoa verrattiin seuraavaksi siten, että laskettiin 'tarkista'- ja 'ongelma'-ryhmien virtakäyrien absoluuttinen erotus 'ok'-ryhmän virtakäyrään. Tällä tavoin saatiin 'tarkista'- ja 'ongelma'-ryhmän virtakäyrien virta-arvon taso parhaiten esille. Erityisesti virta-arvojen tason erotus saatiin paremmin esille sellaisilla virtakäyrillä, missä virtakäyrän amplitudissa oli havaittavissa enemmän vaihtelua. Virtakäyrien absoluuttisen erotuksen avulla saatiin eroteltua 'ok'-, 'tarkista'- ja 'ongelma'-ryhmän virtamittauskäyrien tasoerot toisistaan.

Kuvassa 17 on havainnollistettu yksittäisen virtamittauskäyrän tason erotusta 'ok'-ryhmän tasoon. Kuvassa 17 on vihreällä värillä 'ok'-ryhmän vertailuarvon virtamittauskäyrä ja sinisellä värillä yksittäinen virtamittauskäyrä.

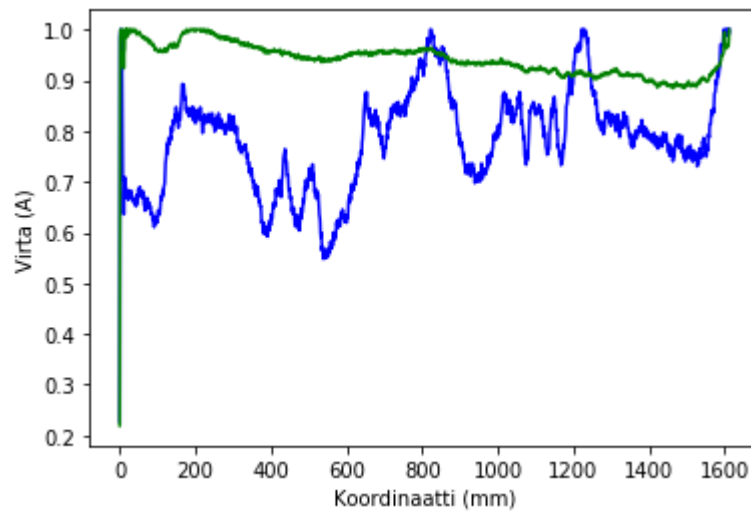


**Kuva 17.** Yksittäisen virtamittauskäyrän tason erotus 'ok'-ryhmän tasoon

Virtakäyrien muotoa ja amplitudin vaihtelevuutta kartoitettiin ensimmäiseksi tarkastelemalla visuaalisesti kaikkien ryhmien virtakäyrien muotoa. Asiantuntijoiden avustuksella tarkasteltiin 'tarkista'- ja 'ongelma'-ryhmään kuuluvien virtakäyrien amplitudin vaihtelevuutta ja muotoa. Virtakäyrän amplitudin ja muodon vaihtelevuus merkitsee, että tasan nopeuden tuottamiseksi tarvittavan voiman määrä vaihtelee pääakselin eri kohdissa. Ideaalinen virtakäyrän muoto on vaakasuora, mutta käytännössä virtakäyrän muodossa on aina vähän vaihtelevuutta. Tehotiheysspektrin maksimiarvolla saatiin hyvin eroteltua sellaiset virtamittauskäyrät, joissa oli amplitudin vaihtelevuutta. Tehotiheysspektrin maksimiarvon lisäksi virtakäyrän keskihajonta ilmaisi virtakäyrän amplitudin vaihtelevuutta. Tehotiheysspektrin ja keskihajonnan avulla saatiin eroteltua 'ok'-, 'tarkista'- ja 'ongelma'-ryhmän virtakäyrien amplitudin vaihtelevuuserot toisistaan.

Kuvassa 18 on havainnollistettu yksittäisen virtamittauskäyrän amplitudin ja muodon vaihtelevuutta 'ok'-ryhmän vertailukäyrään nähden. Kuvassa 18 on vihreällä värillä 'ok'-ryhmän vertailukäyrä ja sinisellä värillä yksittäinen virtakäyrä.





**Kuva 18.** Yksittäisen virtamittauskäyrän amplitudin ja muodon vaihtelu 'ok'-ryhmän vertailukäyrään verrattuna.

Piirteitä tutkittaessa arvioitiin huipukkuus- ja vinousarvoja eri ryhmiin kuuluvista virtamittauskäyristä. Arvioinnissa tuli ilmi, että huipukkuus- vinousarvot menevät päällekkäin eri ryhmien kesken. Päällekkäisyydestä johtuen huipukkuutta ja vinoutta ei käytetty virtamittauskäyrien piirteinä. Eri ryhmien virtamittauskäyristä arvioitiin entropian sopivuutta yhdeksi piirteeksi, mutta eri ryhmien virtamittauskäyrät menivät päällekkäin entropian osalta. Tästä syystä entropiaa ei käytetty virtamittauskäyrien piirteenä.

Edellä esitetyillä piirteillä saatiin parhaiten kaikkien eri levytyökonetyyppien eri pääakselien kolmen eri ryhmän piirteet parhaiten kuvattua. Piirteiden lukumääräksi tuli neljä kappaletta ja valitut piirteet olivat seuraavat:

- 'ok'-ryhmän vertailukäyrän ja yksittäiseen virtamittauskäyrän erotuksen keskiarvojen summa.
- erotuksen keskihajonta.
- yksittäisen pääakselin virtamittauksen keskihajonta.
- yksittäisen pääakselin virtamittauksen tehotiheyspektrin maksimiarvo.

## 5.4 Virtamittaustietojen luokittelu

Piirteiden valinnan jälkeen tehtiin eri levytyökonemallien pääakseleiden luokittelua kolmen eri luokan mukaan. Virtamittausten luokitus tehtiin asiantuntijoiden avustuksella, jolloin saatiin mahdollisimman hyvä luokitustieto virtamittauksille. Luokitteluluokat olivat 'ok'-, 'tarkista'- ja 'ongelma'-luokka. 'Ok'-luokan mukaan määriteltiin jokaiselle eri levytyökonetyypille pääakseleittain 'ok'-luokan vertailuarvo, jonka mukaan tehtiin luokittelu 'ok'- 'tarkista'- ja 'ongelma'-luokkiin.

Jokaisen eri levytyökonetyypin pääakseleiden virtamittaukset luokiteltiin kolmeen eri luokkaan ja muodostettiin eri luokittelijoiden opetus- ja testaustiedot. Käytettävien tiedon perusteella ja päädyttiin neljään eri luokittelijaan, joihin perehdyttiin tarkemmin.

Valitut luokittelijat olivat seuraavat:

- k-lähimmän naapurin luokittelija.
- tukivektorikone.
- päätöspuu-luokittelija.
- lineaarinen diskriminanttianalyysi.

Valituilla luokittelijoilla luokiteltiin eri levytyökonemallin eri pääakseleiden virtamittauksia ja virtamittaustietojen suhteellisen pienen lukumäärän ansiosta luokittelu oli nopeahkoa. Tämän ansiosta voitiin kohtuullisen nopeasti kokeilla yrityksen ja erehdyksen kautta eri parametrien vaikutusta luokittelun tulokseen ja tarkkuuteen. Jokaisen luokittelijan kanssa käytettiin apuna Python-ohjelmointikielessä olevaa automaattista tiedon jakoa opetus- ja testaustietoon, jossa jako tehdään jokaisella kerralla satunnaisesti.

Kahdessa eri levytyökonetyypissä on sama mekaaninen ja sähköinen rakenne yhden pääakselin osalta. Luokittelussa arvioitiin ja testattiin näiden kahden saman rakenteisen pääakselin virtamittaukset myös yhdessä. Tällä tavoin saatiin enemmän virtamittauksia luokiteltavaksi ja jokaiseen eri luokkaan kuuluvia esimerkkejä enemmän. Suuremmalla virtamittausten lukumäärällä pystyttiin paremmin käyttämään hyväksi Python-ohjelmointikielessä olevaa satunnaista opetus- ja testaustiedon jakoa. Tällä tavoin jokaiseen eri luokkaan saatiin opetus- ja testaustietoa jokaiselle testikierrokselle.

Virtamittauksissa mitä ei voitu yhdistellä keskenään suoritettiin myös käsin tehtävää opetus- ja testaustiedon jakoa. Tällä tavoin pyrittiin varmistamaan, että jokaiseen eri luokkaan saataisiin opetus- ja testaustietoa.

## 5.5 Luokittelun arviointi

Kokonaisuutena kaikki neljä käytettyä luokittelijaa luokittelivat 'ongelma'-luokkaan kuuluvat virtamittaukset todella hyvin. Suurimmat ongelmat kohdattiin tukivektorikoneen ja lineaariseen diskriminantti-analyysiin perustuvan luokittelijan kanssa. Muutamisissa 'ongelma'-luokkaan kuuluvissa virtamittauksissa tuli virheellisiä luokituksia tukivektorikonetta ja lineaariseen diskriminantti-analyysiin perustuvaa luokittelijaa käytettäessä. Parhaimmat tulokset saatiin k-lähimmän naapurin-luokittelijalla ja päätöspuu-luokittelijalla. Parhaissa tapauksissa molemmat luokittelijat luokittelivat kaikki 'ongelma'-luokkaan kuuluvat virtamittaukset oikein. Parhaiten kaikkien eri levytyökonetyyppien eri pääakseleiden virtamittaukset luokitteli päätöspuu-luokittelija.

Taulukossa 1 on esitetty kolmen eri levytyökonetyypin kahden eri pääakselin luokittelutarkkuuksia neljällä eri luokittelijalla.

*Taulukko 1. Kolmen eri levytyökonetyypin kahden eri pääakselin luokittelutarkkuudet.*

	<b>K-lähimmän naapurin luokittelijan tarkkuus</b>	<b>Tukivektorikone luokittelijan tarkkuus</b>	<b>Päätöspuu luokittelijan tarkkuus</b>	<b>Lineaarinen diskriminantti analyysi luokittelijan tarkkuus</b>
<b>Lävistävä levytyökone: x-akseli</b>	92%	92%	92%	92%
<b>Lävistävä levytyökone: y-akseli</b>	75%	65%	66%	65%
<b>Yhdistelmä-levytyökone: x-akseli</b>	89%	89%	93%	96%
<b>Yhdistelmä-levytyökone: y-akseli</b>	59%	55%	76%	55%
<b>Lineaari levytyökone: x-akseli</b>	84%	84%	84%	84%
<b>Lineaari levytyökone: y-akseli</b>	86%	73%	86%	86%



Kuvasta 19 havaitaan, että 'ongelma'-luokan virtamittaukset luokitellaan hyvin ja ne erottuvat muista luokista. 'Tarkista'-luokan virtamittauksista havaitaan, että osassa 'tarkista'-luokkaan luokitelluista virtamittausarvoista on hyvin lähellä 'ok'-luokan rajaa. Yksi 'ok'-luokan virtamittaus on taasen hyvin lähellä 'tarkista'-luokan rajaa.

Suurin osa 'ok'-luokan virtamittausarvoista on hyvin yhteneväisiä toisiinsa nähdessä ja niiden luokittelun tulos on hyvin yhtenäinen, joka havaitaan kuvassa 19 olevasta tiiviistä 'ok'-luokan tulosjoukosta. 'Tarkista'-luokan rajojen määrittelyn ongelmat näkyvät kuvassa 19 'tarkista'-luokan hajontana.

## 5.6 Levytyökoneen pääakselin vikaantumisen ennustaminen

Levytyökoneen yksittäisen pääakselin vikaantumisen ennustaminen on virtamittausten avulla haastava tehtävä. Pääakselin mekaanisen sovituksen välyksen lisääntyminen havaitaan virta-arvon pienenemisenä tai joissain tapauksissa virta-arvon kasvamisena. Käytettävissä olleissa virtamittauksissa ei ollut yhtäkään tapausta, joissa yhdenkään levytyökoneen pääakselilla olisi esiintynyt välystä.

Tämän diplomityön aikana ei saatu luotettavaa virtamittausta yhdestäkään pääakselin vikaantumisesta. Pääakseleiden vikaantumisia oli todettu, mutta näistä tilanteista ei ollut virtamittausarvoja käytettävissä. Vikaantumisen ennustamisessa tullaan käyttämään valittujen piirteiden muutoksen seuranta arvioitaessa pääakselin kuluneisuutta. Seuranta kohdistuu pitkälle ajanjaksolle, koska levytyökoneen pääakselin vikaantumiset eivät ole usein toistuvia tapahtumia.

Levytyökoneen valmistumishetkestä pääakselin vikaantumiseen kuluva aika tullaan seuraamaan ja tarkastelemaan. Tällä tavoin saadaan kerättyä tietoa siitä, vaikuttaako virtamittauksessa huomattu mekaanisen sovituksen laatu keskimääräiseen pääakselin vikaantumisaikaan. Tällä tavoin saadaan kerättyä tietoa siitä, miten suuri poikkeama virtamittauksessa aiheuttaa keskimääräisen pääakselin vikaantumisaajan lyhentymisen.

## 5.7 Eri levytyökonemallien vastaavuudet

Kahdessa eri levytyökoneetyypissä on samanlainen mekaaninen ja sähköinen toteutus yhden pääakselin osalta, joten virtamittaustulosten tulisi olla toisiaan vastaavat. Virtamittausten vastaavuutta lähdettiin selvittämään korrelaation avulla. Kahdesta eri levytyökoneetyypistä otettiin saman pääakselin virtamittaukset ja laskettiin korrelaatiokerroin niiden välillä. Virtamittausten lukumäärä oli suhteellisen pieni, joten voitiin laskea kaikkien mekaanisesti ja sähköisesti vastaavien levytyökoneiden pääakselin korrelaatiokerroimet. Korrelaatiokerroimen ollessa riittävän suuri tai pieni, tarkasteltiin myös virtamittausten kuvaajia kyseessä olleelta vertailuparilta. Toinen käytetty menetelmä vastaavuus-

den arviointiin oli liukuvan ikkunan avulla laskettu virtamittauksen keskiarvo. Taulukossa 2 on levytyökone-pareittain korrelaatiokertoimen ja liukuvan keskiarvon tulokset.

**Taulukko 2.** Levytyökone-parit vertailtuna korrelaatiokertoimen ja liukuvan keskiarvon mukaan.

Levytyökone parit	Korrelaatiokerroin	Levytyökoneen pääakselin virran keskiarvo 1	Levytyökoneen pääakselin virran keskiarvo 2
1	0,87	0,07A	0,09A
2	0,64	0,05A	0,07A
3	0,66	0,06A	0,05A
4	0,79	0,03A	0,05A
5	0,35	0,03A	0,24A
6	0,88	0,05A	0,05A
7	0,85	0,07A	0,03A
8	0,63	0,03A	0,10A
9	0,95	0,06A	0,06A
10	0,72	0,05A	0,17A
11	0,89	0,06A	0,11A
12	0,86	0,04A	0,08A
13	0,82	0,04A	0,06A
14	0,93	0,05A	0,05A
15	0,38	0,05A	0,06A
16	0,97	0,03A	0,04A
17	0,69	0,03A	0,09A
18	0,73	0,06A	0,04A
19	0,77	0,07A	0,05A
20	0,85	0,07A	0,05A
21	0,63	0,09A	0,09A
22	0,65	0,15A	0,12A
23	0,65	0,06A	0,11A
24	0,41	0,06A	0,17A
25	0,80	0,05A	0,10A
26	0,80	0,06A	0,10A
27	0,72	0,07A	0,11A
28	0,30	0,12A	0,09A
29	0,41	0,07A	0,15A

Kaikkien 'ok'-luokkaan kuuluvien levytyökoneiden vastaavuudet saatiin ilmaistua korrelaatiokertoimen ja liukuvan keskiarvon avulla. Sellainen levytyökone, joka ei kuulunut 'ok'-luokkaan poikkesi, joko korrelaatiokertoimen tai liukuvan keskiarvon osalta. Kahden eri levytyökonetyypin pääakselin sama mekaaninen ratkaisu pystytettiin havaitsemaan virtamittausarvoista korrelaatiokertoimen ja liukuvan keskiarvon avulla.

Taulukossa 2 viidennessä parissa toinen vertailtavista koneista oli virtamittausarvon luokittelun mukaan 'tarkista'-luokassa ja toinen 'ok'-luokassa. Parissa 5 vertailtavien virtamittauskäyrien amplitudin muodot poikkesivat toisistaan merkittävästi. Vastavasti parissa 10 olivat virtamittauskäyrien muodot yhteneväisiä, mutta toinen pääakseleista oli 'ok'-luokan alarajalla ja toinen 'ok'-luokan ylärajalla. Parit 15, 24, 28 ja 29 poikkesivat toisistaan virtamittauskäyrän amplitudin vaihtelevuuden ja muodon perusteella.

Kahden eri levytyökonetyypin vertailussa havaittiin, että korrelaatiokerroin ilmaisee testiaineiston perusteella eroavaisuudet 'ok'-luokan ja 'ongelma'-luokan virtamittauskäyrien välillä. Havaintoon perustuen kahden eri levytyökonetyypin vertailussa käytettyä korrelaatiokerrointa voitaneen käyttää yhtenä luokittelun piirteenä tulevaisuudessa. Korrelaatiokertoimen käyttäminen piirteenä tarvitsee kuitenkin yhden yksittäisen virtamittauksen referenssiarvoksi, mihin muita virtamittausarvoja verrataan. Tämän diplomityön aikana ei saatu asiantuntijoiden avustuksella määriteltyä yhtä yksittäistä virtamittausta kuvaamaan pääakselin hyväksyttävää referenssiarvoa.

## 6 JOHTOPÄÄTÖKSET

Tässä diplomityössä tutkittiin levytyökoneista pääakselikohtaisesti saatavan virtamittaustiedon analysoinnin ja koneellisen luokittelun eri menetelmiä. Eri menetelmiä vertailtiin keskenään ja diplomityön tuloksena saatiin valittua kaksi parhaiten virtamittaustiedon luokitteluun soveltuvaa luokittelijaa. Parhaiten virtamittaustietojen luokitteluun soveltuviksi todettiin k-lähimmän naapurin luokittelija ja päätöspuu-luokittelija. Alla olevissa kappaleissa on tarkemmin kuvattuna työssä esiin tulleet havainnot ja johtopäätökset.

### 6.1 Yhteenveto

Virtamittauksien analysoinnilla ja koneellisella luokittelulla tunnistettiin virtamittausprosessissa oleva ongelma yhden pääakselin osalta. Levytyökoneiden pääakselien virtamittauksien analysoinnilla ja koneellisella luokittelulla pystytään tunnistamaan luotettavasti 'ok'-luokkaan ja 'ongelma'-luokkaan kuuluvat virtamittaukset. Haasteellisin luokka on 'tarkista'-luokan virtamittaukset, koska niissä osa piirteistä limittyi, joko 'ok-luokan' tai 'ongelma'-luokan kanssa. Koneellisella luokittelulla saavutetaan 'ongelma'-luokan automaattinen tunnistaminen ja tarvittavan informaation toimittaminen asiantuntijoiden tarkastelua ja ratkaisua varten.

Levytyökoneiden pääakseleiden mekaanisen sovituksen laatua saadaan parannettua virtamittauksien analysoinnilla ja koneellisella luokittelulla. Analysointi ja koneellinen luokittelu antaa levytyökoneen valmistuksen loppuvaiheessa tiedon siitä, mihin luokkaan pääakseleiden virtamittaukset kuuluvat. Tällä tavoin havaitaan kaikki 'ongelma'-luokkaan kuuluvat levytyökoneet ja voidaan suorittaa tarvittavat toimenpiteet ennen koneen luovutusta asiakkaalle.

Olemassa olevan levytyökone kannan pääakseleiden virtamittausten analysointi ja koneellinen luokittelu parantaa huollon ennakoitavuutta. Analysoinnilla ja luokittelulla saavutetaan pääakseleiden muutoksien automaattinen tunnistaminen ja raportointi. Tällä tavoin saavutetaan parempi tieto siitä, miten eri levytyökoneet ovat kuluneet eri pääakseleiden osalta. Pääakseleiden kuluneisuuden perusteella voidaan tehdä tarkempia suunnitelmia tarvittavien huoltojen ajankohdista ja laajuuksista.



## 6.2 Sovellettavuus

Kehitetyjä virtamittausten analysoinnin algoritmeja ja koneellisen luokittelun menetelmiä voidaan hyödyntää esimerkiksi levytyökone valmistajan pilvipalvelun osana. Pilvipalvelussa tapahtuva automaattinen analysointi ja luokittelu tehostaa olemassa olevan kone kannan huollon suunnittelua ja ennakoitavuutta. Kehitetyt menetelmät voidaan integroida virtamittausohjelmiston osaksi ja tällä tavoin saadaan virtamittauksen päätteeksi analyysi ja luokittelun tulos nähtäville. Tämä tehostaa levytyökoneiden valmistuksen loppuvaiheessa tehtäviä virtamittauksia ja niiden analyysseja.

## 6.3 Jatkotoimenpiteet

Yhdessä pääakselin virtamittausprosessissa oleva virhe pitää poistaa, koska muuten virtamittauksesta saatavat tulokset eivät ole luotettavia. Tässä tapauksessa virtamittaukseen pääsee vaikuttamaan ulkoinen muuttuja, joka on säädettävissä. Virtamittausten tarkempaan validointiin pitää kiinnittää huomiota. Tässä diplomityössä havaittiin useita epätodellisia virtamittauksia, jotka olivat puutteellisia tiedon osalta. Osasta puuttui tietoa ja osa oli jäänyt kesken, myös liian pitkiä virtamittautietoja esiintyi.

Olemassa olevasta levytyökone kannasta tehtävää virtamittausta pitää suorittaa järjestelmällisemmin. Levytyökoneiden pääakseleiden virtamittaushistoriassa on puutteita ja se aiheuttaa ongelmia muutosten seurantaan. Virtamittaushistoria pitäisi saada riittävän kattavaksi, että pystyttäisiin tunnistamaan muutoskohdat paremmin.

Virtamittausten luokittelua tulee vielä tarkentaa, koska 'tarkista'-luokkaan kuuluvia virtamittautietoja on hyvin lähellä 'ongelma'- tai 'ok'-luokkaa. Pitäisi pystyä tarkemmin rajaamaan eri luokat toisistaan. Käytettäviä piirteitä tulee arvioida uudelleen, kun saadaan tarkempaa tietoa analysoinnin ja luokittelun toimivuudesta ajan kuluessa. Käytettävät luokittelijat tulee arvioida uudelleen ajan kuluessa ja arvioida tarvitaanko levytyökonetyyppi- ja pääakselikohtaisesti eri luokittelijat.

## LÄHTEET

- [1] Peter J. Brockwell, Richard A. Davis, Introduction to Time Series and Forecasting, Second Edition, Springer , New York 2002, s.429
- [2] Robert H. Shumway, David S. Stoffer, Time Series Analysis and Its Applications, With R Examples Third Edition, Springer , New York 2011, s.591
- [3] Nist Sematech, Engineering statistics handbook [WWW] [Viitattu 2.6.2018]. Saatavissa:  
<https://www.itl.nist.gov/div898/handbook/eda/section3/eda35b.htm>
- [4] Ronald E. Walpole, Raymond H. Myers, Sharon L. Myers, Keying Ye. Probability & Statistics for Engineers & Scientists, Ninth Edition, Pearson , Boston 2012, s.785
- [5] Nist Sematech, Engineering statistics handbook [WWW] [Viitattu 4.6.2018]. Saatavissa:  
<https://www.itl.nist.gov/div898/handbook/eda/section3/eda33e.htm>
- [6] Wikipedia. [WWW] [Viitattu 8.6.2018]. Saatavissa:  
[https://en.wikipedia.org/wiki/Correlation\\_and\\_dependence](https://en.wikipedia.org/wiki/Correlation_and_dependence)
- [7] Ilkka Mellin. Aalto University, Department of Mathematics and System Analysis [WWW] [Viitattu 8.6.2018]. Saatavissa:  
<http://math.tkk.fi/opetus/sovtoda/luennot/vanhat/TILRI100.pdf>
- [8] Statistics How To. [WWW] [Viitattu 8.6.2018]. Saatavissa:  
<http://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/#Pearson>
- [9] Monica Borda. Fundamentals in Information Theory and Coding, Springer, Berlin 2011, s.388

- [10] Wikipedia. [WWW] [Viitattu 11.6.2018]. Saatavissa:  
[https://en.wikipedia.org/wiki/Entropy\\_\(information\\_theory\)](https://en.wikipedia.org/wiki/Entropy_(information_theory))
- [11] University of Helsinki. [WWW] [Viitattu 12.6.2018]. Saatavissa:  
[http://www.courses.physics.helsinki.fi/astro/havaitsevaII/Radiomoniste/liite\\_b.pdf](http://www.courses.physics.helsinki.fi/astro/havaitsevaII/Radiomoniste/liite_b.pdf)
- [12] John G. Proakis, Dimitris G. Manolakis. Digital Signal Processing, Third Edition, Prentice-Hall, Inc, New Jersey 1996, s.960
- [13] Shai Shalev-Scwartz, Shai Ben-David. Understanding Machine Learning, From Theory to Algorithms, Cambridge University Press, New York 2014, s.437
- [14] Alex Smola and S.V.N. Vishwanathan. Introduction to Machine Learning, Cambridge University Press, United Kingdom 2008, s.219
- [15] Peter Harrington. Machine Learning in Action, Manning Publications Co, Shelter Island, NY 2012, s.347
- [16] Soroush Nasiriany, Garret Thomas, Walliam Wang, Alex Yang. A Comprehensive Guide to Machine Learning. University of California, Berkeley [WWW] [Viitattu 1.7.2018]. Saatavilla:  
<http://snasiriany.me/files/ml-book.pdf>
- [17] Sebastian Raschka. Python Machine Learning, Packt Publishing Ltd, Birmingham B3 2PB, UK 2015, s.417
- [18] Christopher M. Bishop. Pattern Recognition And Machine Learning, Springer, Singapore 2006, s.729